Boosting to the Top: Measurements of Boosted Top Quarks and Higgs Bosons with the ATLAS Detector at the Large Hadron Collider

Michael Fenton

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

July 2019

Abstract

Between 2015 and 2018, proton-proton collisions were performed at the highest energy ever achieved in man-made particle accelerators, with the Large Hadron Collider at CERN. This thesis presents measurements performed with the ATLAS detector which fully exploit this energy, of the high momentum production of the most massive fundamental particles known to exist: the top quark and the Higgs boson. Firstly, the differential cross-section of highly boosted top quarks, measured in the lepton+jets channel using the 3.2 fb^{-1} of $\sqrt{s}$ = 13 TeV collected in 2015, is presented. Two kinematic properties of the boosted top quark, $p_T^{\mathrm{t,had}}$ and $\left|y^{\mathrm{t,had}}\right|$, are unfolded to a fiducial volume and compared to a number of state-of-the-art Monte Carlo simulations. Then, using 36 fb^{-1} of data from 2015 and 2016, the search for associated production of a top quark pair with a Higgs boson decaying to a pair of bottom quarks is discussed. Particular emphasis is placed on the high p_T boosted channel, the first time such an analysis has been published by the ATLAS collaboration. A limit of 2.0 times the Standard Model expectation is set on $t\bar{t}H, H \to b\bar{b}$ production alone, which after combining with analyses in other channels leads to $> 5\sigma$ observation of both $t\bar{t}H$ production and $H \to b\bar{b}$ decay. Finally, a number of optimisations of the boosted $t\bar{t}H, H \to b\bar{b}$ analysis are presented, in view of a future publication of the combined $t\bar{t}H, H \to b\bar{b}$ search using the full 140 fb^{-1} of data collected by the end of 2018.

Contents

Abstract **i**

Acknowledgements **xiii**

Declaration **xv**

1 Introduction **1**

2 The Standard Model of Particle Physics **3**

2.1 $\mathcal{L}_{\text{Gauge}}$: The Fundamental Forces of Nature 6

2.2 $\mathcal{L}_{\text{Fermions}}$: Quarks and Leptons 7

2.2.1 The Top Quark 8

2.3 $\mathcal{L}_{\text{Higgs}}$: Electroweak Symmetry Breaking and the Higgs Mechanism 9

2.3.1 The Higgs Boson 11

2.4 $\mathcal{L}_{\text{Yukawa}}$: Acquiring Fermion Masses 14

2.4.1 The Top-Higgs Coupling 15

2.5 $\mathcal{L}_{\text{BSM}}$: Beyond the Standard Model 15

3 Analysis Techniques **17**

3.1 Boosted Particles 17

3.2 Unfolding 18

3.3 Profile Likelihood Fit 20

3.4 Multivariate Analysis 22

3.4.1 Common Concepts 23

3.4.2 Boosted Decision Trees 24

3.4.3 Deep Neural Networks 26

4 ATLAS and the LHC **31**

4.1 The Large Hadron Collider 31

4.1.1 Operation 32

4.2 ATLAS 36

4.2.1 Inner Detector 37

4.2.2 Calorimeters . . . 40
4.2.3 Muon Spectrometer . . . 42
4.2.4 Magnets . . . 44
4.2.5 Trigger . . . 46
4.2.6 FTK Integration in Muon Trigger Chain . . . 48

5 Object Definitions **57**
5.1 Low Level Objects . . . 57
5.1.1 Clusters and Towers . . . 57
5.1.2 Tracks . . . 58
5.2 Leptons . . . 60
5.2.1 Electrons . . . 60
5.2.2 Muons . . . 62
5.3 Jets . . . 63
5.3.1 Small Jets . . . 64
5.3.2 B-Tagging . . . 66
5.3.3 Track Jets . . . 71
5.3.4 Large Jets . . . 71
5.3.5 Jet Substructure . . . 72
5.3.6 Large Jet Top-tagging . . . 74
5.3.7 Reclustered Jets . . . 76
5.4 Overlap Removal . . . 76
5.5 Missing Transverse Momentum . . . 77

6 Monte-Carlo Simulation **78**
6.1 Factorisation Theorem . . . 78
6.2 Parton Distribution Functions . . . 80
6.3 Matrix Element Generation . . . 80
6.4 Parton Shower . . . 81
6.4.1 Enhancing $g \to b\bar{b}$ Splitting in Pythia8 . . . 82
6.5 Hadronisation . . . 86
6.6 Underlying Event / Multi-parton Interactions . . . 86
6.7 Detector Simulation . . . 87
6.8 Generated Samples . . . 87

7 Differential Cross-Section of Boosted Top Quarks **91**
7.1 Event Selection . . . 91
7.1.1 Detector Level Selection . . . 92
7.1.2 Particle Level Selection . . . 92

7.2 Background Estimation . . . 95
7.2.1 Data Driven W+Jets Scale Factors . . . 95
7.2.2 Data Driven QCD Multijet Estimate . . . 96
7.2.3 Control Plots . . . 97
7.3 Unfolding Tests . . . 101
7.4 Systematic Uncertainties . . . 104
7.5 Results . . . 107
7.5.1 Conclusions . . . 113

8 Searching for $t\bar{t}H, H \to b\bar{b}$ using Boosted Objects 114
8.1 Event Selection and Categorisation . . . 115
8.1.1 Resolved Dilepton . . . 115
8.1.2 Resolved Lepton+Jets . . . 115
8.1.3 Boosted Lepton+Jets . . . 118
8.2 Background Modelling . . . 119
8.3 Multivariate Strategy . . . 121
8.4 Systematic Uncertainties . . . 123
8.5 Results . . . 125
8.6 Full $t\bar{t}H$ Combination . . . 129

9 Optimising the Boosted $t\bar{t}H, H \to b\bar{b}$ Channel 133
9.1 Track Jet b-Tagging . . . 133
9.1.1 Track Jet Performance . . . 134
9.1.2 High Purity Top-tag Region . . . 141
9.1.3 Trimmed Top + Higgs Region . . . 143
9.1.4 Track Jets Conclusions . . . 144
9.2 Reclustered Jet Substructure . . . 145
9.2.1 Proxy Jet . . . 145
9.2.2 Systematic Uncertainties . . . 147
9.2.3 Application to $t\bar{t}H, H \to b\bar{b}$ . . . 155
9.3 Multiclass Jet Tagging using Deep Neural Networks . . . 157
9.3.1 DNN Training . . . 157
9.3.2 Selecting Events using DNN Tagger . . . 164
9.3.3 Classifier Training in DNN Region . . . 170
9.3.4 Classifier Training in Baseline Region . . . 172
9.3.5 Fits . . . 174
9.3.6 DNN Tagging Summary . . . 182
9.4 Outlook . . . 183

10 Conclusions **184**

List of Figures

2.1 Summary plot of SM cross-section measurements, over many orders of magnitude and at varying centre-of-mass energies, as measured by ATLAS. 4
2.2 The SM Lagrangrian written on the side of a coffee cup. 5
2.3 The particle content of the Standard Model of Particle Physics. 6
2.4 Leading order Feynman diagrams for $t\bar{t}$ production at the LHC. 9
2.5 Breakdown of top quark final states. 10
2.6 Graphical representation of the Higgs potential. 11
2.7 Representative Feynman diagrams for the main Higgs boson production modes at the LHC. 12
2.8 Higgs boson production cross-sections in the SM, as a function of centre-of-mass energy. 13
2.9 Branching ratios for the SM Higgs boson. 14

3.1 Cartoon depiction of the transition between the resolved and boosted regimes, for the case of a hadronically decaying top quark. 18
3.2 Example schematic of a Decision Tree. 24
3.3 Model of a Deep Neural Network containing two hidden layers. 27
3.4 A toy example of a forward pass for a simple, fully connected network. 28
3.5 Examples of common non-linear activation functions. 28

4.1 Diagram showing the CERN accelerator complex. 33
4.2 The luminosity delivered by the LHC and recorded by ATLAS through each year of Run 2 data taking. 34
4.3 The total luminosity delivered and recorded over the whole of Run 2 by the ATLAS detector. 35
4.4 The average number of visible pp interactions per bunch crossing in ATLAS, per year and total. 36
4.5 The ATLAS Detector. 37
4.6 Cartoon demonstrating how different types of particle interact with each layer of the ATLAS detector. 38
4.7 The layout of the ATLAS inner detector. 39

4.8 The ATLAS calorimeter system. 41
4.9 The ATLAS muon and magnet systems. 43
4.10 Schematic view of the muon spectrometer. 43
4.11 The geometry of the ATLAS magnet system. 45
4.12 Schematic of the ATLAS TDAQ system, as used in Run 2. 46
4.13 FTK tracking efficiency relative to standard offline tracking in ATLAS, for both muons and pions. 48
4.14 Flow chart depicting the single muon trigger chain used in ATLAS. 49
4.15 2D plot of the Muon candidate ID track p_T vs the removed track p_T in the FTK isolation algorithm. 51
4.16 The type 1 and type 2 efficiencies for the isolation algorithms. 53
4.17 The tag and probe efficiencies for the standard single muon chain and the FTK isolation chain. 54
4.18 The control distributions for the precision ID tracking algorithm, both with and without the FTK based isolation in the chain. 55
4.19 The CPU processing time per event in enhanced bias data, for both the standard chain and FTK chain. 56

5.1 Jet finding results from an example event of the various jet algorithms described. 64
5.2 Cartoon summary of calibration procedure for small jets. 65
5.3 The results of the in situ calibrations as a function of jet p_T. 67
5.4 The full uncertainty on the JES as a function of both p_T and η. 67
5.5 Diagram demonstrating the characteristic displaced vertex in a small jet originating from a B-hadron. 68
5.6 The rejection rates for light jets and charm jets, for different configurations of the MV2 algorithm. 70
5.7 The requirements applied for the top-tagging algorithm as a function of jet p_T, for mass and τ_{32}. 75
5.8 The tagging efficiency and background rejection for top quarks as a function of jet p_T. 75

6.1 Cartoon depiction of the factorised stages in MC generation. 79
6.2 Example Feynman diagram for $t\bar{t} + b\bar{b}$ production. 82
6.3 Distributions of the number of b-jets in the generated $t\bar{t}$ samples, with and without event weights applied. 84
6.4 Kinematic distributions of the b-jets in the generated $t\bar{t}$ samples with event weights applied. 84
6.5 Kinematic distributions of the $t\bar{t}$ system in the generated $t\bar{t}$ samples with event weights applied. 85

7.1 Diagram summarising the topology and event selection of the $t\bar{t}$ cross-section analysis. . . . 92
7.2 The unfolding acceptance correction factors. . . . 93
7.3 The unfolding efficiency correction factors. . . . 94
7.4 The unfolding migration matrices. . . . 94
7.5 Detector level plots of the number of large jets and the large jet p_T. . . . 98
7.6 Detector level plots of the lepton p_T and η. . . . 98
7.7 Detector level plots of E_T^{miss} and m_T^W. . . . 99
7.8 Detector level plots of the hadronic top candidate jet τ_{32} and mass. . . . 99
7.9 Detector level plots of the hadronic top candidate jet kinematics. . . . 100
7.10 The $p_T^{\text{t,had}}$ and $\left|y^{\text{t,had}}\right|$ closure tests. . . . 102
7.11 The p_T dependent stress tests. . . . 102
7.12 The $y^{t\bar{t}}$ dependent stress tests. . . . 103
7.13 Plots showing the largest fractional contributions to the uncertainty of the absolute differential cross-section, after unfolding. . . . 104
7.14 Plots showing the largest fractional contributions to the uncertainty of the relative differential cross-section, after unfolding. . . . 105
7.15 The m_T^W distributions in a 0 top-tag, 0 b-tag validation region. . . . 106
7.16 Normalised differential cross-sections in the boosted topology, as a function of top quark transverse momentum. . . . 107
7.17 Normalised differential cross-sections in the boosted topology, as a function of top quark absolute rapidity. . . . 108
7.18 Absolute differential cross-sections in the boosted topology, as a function of top quark transverse momentum. . . . 109
7.19 Absolute differential cross-sections in the boosted topology, as a function of top quark absolute rapidity. . . . 110

8.1 Event Categorisation based on PCB scores in the dilepton $t\bar{t}H, H \to b\bar{b}$ channel. 116
8.2 Event Categorisation based on PCB scores in the resolved lepton+jets $t\bar{t}H, H \to b\bar{b}$ channel. . . . 117
8.3 Cartoon summary of the boosted selection in the $t\bar{t}H, H \to b\bar{b}$ analysis. . . . 118
8.4 The relative background fractions in each region of the $t\bar{t}H, H \to b\bar{b}$ analysis. . 119
8.5 The relative fractions of the various $t\bar{t}$+b sub-components in the $t\bar{t}H, H \to b\bar{b}$ analysis. . . . 120
8.6 The Higgs candidate mass distributions in the boosted signal region and the tightest resolved signal region of the $t\bar{t}H, H \to b\bar{b}$ analysis. . . . 121
8.7 The BDT discriminant distribution in the boosted $t\bar{t}H, H \to b\bar{b}$ signal region, pre and post fit. . . . 123

8.8 The observed number of events in the $t\bar{t}H, H \to b\bar{b}$ analysis as a function of $\log_{10} S/B$. 126
8.9 The pulls and constraints of the 20 largest systematic uncertainties post-fit, in the $t\bar{t}H, H \to b\bar{b}$ analysis. 128
8.10 The observed number of events in the combined $\sqrt{s} = 13$ TeV $t\bar{t}H$ analyses. . . 130
8.11 The measured signal strengths in each individual analysis, as well as the full $\sqrt{s} = 13$ TeV $t\bar{t}H$ combination. 131

9.1 Binned and integrated tagging efficiencies for b-jets, as a function of small jet p_T and top jet p_T, in a $t\bar{t}$ MC sample. 135
9.2 Binned and integrated tagging efficiencies for b-jets within $\Delta R < 1.0$ of the top jet, as a function of small jet p_T and top jet p_T, in a $t\bar{t}$ MC sample. 136
9.3 Binned and integrated rejection rates for charm jets, as a function of small jet p_T and top jet p_T, in a $t\bar{t}$ MC sample. 137
9.4 Binned and integrated rejection rates for charm jets within $\Delta R < 1.0$ of the top jet, as a function of small jet p_T and top jet p_T, in a $t\bar{t}$ MC sample. 138
9.5 Binned and integrated rejection rates for light jets, as a function of small jet p_T and top jet p_T, in a $t\bar{t}$ MC sample. 139
9.6 Integrated and binned rejection rates for light jets within $\Delta R < 1.0$ of the top jet, as a function of small jet p_T and top jet p_T, in a $t\bar{t}$ MC sample. 140
9.7 Cartoon summary of the high purity top-tag region. 142
9.8 Cartoon summary of the trimmed top+Higgs region. 143
9.9 The difference between the reclustered jet and the proxy jet for each component of the jet 4-vector, normalised by the reclustered jet 4-vector. 146
9.10 Distributions of r_{track} and R_{track} for the six JSS variables. 148
9.11 The final systematic uncertainties for the τ_{21} variable, in four bins of m/p_T. . . 149
9.12 The final systematic uncertainties for the τ_{32} variable, in four bins of m/p_T. . . 150
9.13 The final systematic uncertainties for the $\sqrt{d_{12}}$ variable, in four bins of m/p_T. . 151
9.14 The final systematic uncertainties for the $\sqrt{d_{23}}$ variable, in four bins of m/p_T. . 152
9.15 The final systematic uncertainties for the D_2 variable, in four bins of m/p_T. . . 153
9.16 The final systematic uncertainties for the Q_W variable, in four bins of m/p_T. . . 154
9.17 Separation plots for the six reclustered JSS variables, for the Higgs jet in the boosted $t\bar{t}H, H \to b\bar{b}$ selection. 156
9.18 The separation for each input variable used in the DNN training. 159
9.19 Three of the input variables for the DNN; the reclustered jet mass, the invariant mass of the b-tagged subjets, and the sub-leading (in MV2) subjet p_T. 160
9.20 Confusion matrix for the DNN output. 161
9.21 DNN output distributions. 163
9.22 The Higgs jet mass distribution, in the baseline region and the DNN region. . . 165

9.23 The output DNN distributions for the selected top jet in the DNN region. 166
9.24 Kinematic distributions for the selected top jet in the DNN region. 167
9.25 The overlap with the resolved regions in the DNN region. 168
9.26 The overlap with the resolved regions in the baseline region. 169
9.27 The Higgs jet Higgs score and top jet top score, used in the DNN region classification BDT training. 170
9.28 Separation plot for the classification BDT in the DNN region. 172
9.29 The DNN Higgs and top scores for the Higgs candidate and the top candidate in the baseline region. 173
9.30 Separation plots for the retrained baseline and baseline+DNN regions. 174
9.31 The data and MC distributions for the retrained baseline BDT, the baseline+DNN BDT, and the DNN region BDT. 175
9.32 Plots demonstrating the effect of the matrix element generator systematic in the retrained baseline, the baseline+DNN, and the DNN regions. 177
9.33 Plots demonstrating the effect of the first light mistagging component from the b-tagging eigenvalues, in the retrained baseline, the baseline+DNN, and the DNN regions. 178
9.34 The distribution of the error on μ ($\Delta\mu$) in the three systematic models, for the retrained baseline, the baseline+DNN, and the DNN regions. 180
9.35 The pulls and constraints of the 20 largest systematic uncertainties in the instrumental fits. 181

List of Tables

6.1 The cross-section and corresponding statistical errors for the $t\bar{t}$ + $b\bar{b}$ process, in both dilepton and lepton+jets selections, of 1 million inclusive $t\bar{t}$ events generated with various enhancement settings. 85

7.1 Event yields after the detector level selection of the $t\bar{t}$ cross-section analysis. . . 93
7.2 Comparisons of total fiducial cross-section in data and several MC generator set ups. 111
7.3 Summary of χ^2 and p-values for the absolute distributions. 112
7.4 Summary of χ^2 and p-values for the relative distributions. 112
7.5 Correlation matrix for the normalised p_T distribution. 113

8.1 Input variables for the classification BDT in the boosted $t\bar{t}H, H \to b\bar{b}$ signal region. 122
8.2 TMVA BDT hyperparameters used for the classification BDT in the boosted region. 122
8.3 Breakdown of the sources of systematic uncertainty in the $t\bar{t}H, H \to b\bar{b}$ analysis. 124
8.4 The sources of systematic uncertainty on the $t\bar{t}$+jets background in the $t\bar{t}H, H \to b\bar{b}$ analysis. 125
8.5 Signal strengths measured in Asimov data for the four scenarios of how to treat boosted events in the $t\bar{t}H, H \to b\bar{b}$ analysis. 125
8.6 The post-fit uncertainties in the $t\bar{t}H, H \to b\bar{b}$ analysis. 127
8.7 Summary of the impact of each group of systematic uncertainties in the Run 2 $t\bar{t}H$ combination. 129
8.8 Summary of the $\sqrt{s} = 13$ TeV analyses and combination as well as the Run 1+Run 2 $t\bar{t}H$ combination. 132

9.1 Yields in the high purity top-tag region, comparing the different jet collections used for b-tagging. 142
9.2 Yields in the trimmed top+Higgs region. 144
9.3 List of variables included in DNN training. 158
9.4 The hyperparameters used in the DNN training. 162
9.5 Yields in the baseline and DNN selections, based on 140 fb^{-1} of MC. 164
9.6 Input variables used for the classification BDT in the DNN region. 171
9.7 Hyperparameters for the classification BDT training in the DNN region. 171

9.8 Results from the 140 fb^{-1}, boosted only fits to the three analysis options. 179

Acknowledgements

The existence of this document is predicated on the contributions of a great, great many people, to whom it is be impossible to sufficiently acknowledge and thank in this small space.

Firstly I would like to thank my supervisor Mark Owen, without whom I could not have completed this thesis. I am aware of a great many supervisor-student pairs covering the full spectrum of possible support and knowledge transfer, and I can confidently say that Mark has been the best supervisor out there. I am also thankful to my second supervisor Tony Doyle, who first introduced me to the Glasgow PPE group, and who has been a constant source of support and encouragement. I am also forever indebted to the help and support of Andrea Knue, without whom I probably would not be doing this PhD in the first place, and Sarah Boutle, who was constantly there to help me in every aspect. I certainly couldn't have done it without them, though at least I probably would have drank less beer. On that note, I am also grateful to Chris, Christoph, Ian, Federica, John, and every other member of the Glasgow group, who made my time in the Kelvin so enjoyable. Chloe and Murdo were the best teammates I could ask for.

My friends in Glasgow, including the Twerking Pandas Niall, Liam, Livvy and Courtney; my tramp teammates Ben, Tony, Jon, Rosie, Jamie, Ceri; and my old flatmates Linus, Aaron, Neil and Dave; I'd probably be a lot healthier but have a lot less fun tales to tell without you all.

I also had the great fortune to meet many amazing people in my two years in Geneva. Special mention must go to my cooking mentor Bobby, my Smash bro Duncan, my partner in crime Adam, and my wisest friend Xanthe. Venting about $t\bar{t}H$ over coffee with Jennet kept me sane. Jake and Bijan are in a constant tug of war over the title of "Best Human I Know". The Universe is behind the door, where it belongs.

The Department of Mysteries, including Will, Liam, Cameron, Ellis, Anna, Tim, Harry, Tom, Rose, Iwan: our day trips to Geneva and night trips to L'usine were probably the highlight of my time in Europe, and I wouldn't sacrifice them for anything. I enjoyed many mountain adventures and après skis with Lewis, Chris, Jenny, Joe, Amal, Yallup, Sun, and the Woodland Critters, and many a Wednesday Beer&Kebab with Stoyan, David, Shane, Eddie, Charlie, Emmy, Deshan, Vas, Alice, Nelson, Lauren, Jay, and a great many more.

Finally, and most importantly of all, I would like to thank my family, for their unending support throughout my life and in particular during the last 4 years. Mum and Dad, you have always done far too much for me, and I am forever grateful to you both. I also want to thank

my sister Sally for her loving support and friendship; I've *mostly* gotten over all the times you used to hide my toys or push me into ponds. I am also extremely grateful to have two of the best grandparents in the world still by my side. To Granny Christa, *danke schön.* To Grotty Grandad, *slàinte*. Dedicated to Grandad Paddy and Granny Alie.

I couldn't have done it without you all.

Declaration

The material presented in this thesis is the result of my own work in the Experimental Particle Physics group of the School of Physics and Astronomy at the University of Glasgow. It has not been submitted for any other degree at the University of Glasgow or any other institution.

The work was performed as part of the ATLAS Collaboration at the LHC and as such was made possible only via collaboration with many others. I contributed to the experimental operation and data collection of ATLAS via control room shifts, during which time I was responsible for the trigger operation in both data taking and calibration runs. Chapters 1-6 contain mostly introductory material, relating to the underlying theoretical framework and experimental apparatus with which the work was performed. Section 4.2.6 further contains work I performed as part of my qualification to ATLAS authorship, implementing FTK tracking information into single muon trigger chains. Section 6.4.1 also contains a study I performed into Sudakov enhancement of particular processes for the purpose of enhanced statistics in particular regions of phase space. Chapter 7 is a summary of the boosted analysis from Ref [1], on which I was the lead analyst. Chapter 8 summarises Ref [2], on which I contributed studies on MVA training, fitting, validation regions, and cross-checks. Chapter 9 presents a number of studies I performed as the lead analyst in the boosted channel for a forthcoming $t\bar{t}H$ publication.

Chapter 1

Introduction

The story so far: In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move.

– Douglas Adams

Often in the three and a half years in which I have been performing the work presented in this document, I have been asked the question: why did you choose to study particle physics? Why is it interesting? Why *bother?* The answer is that I like puzzles, of all sorts. And particle physics is the grandest puzzle of all: that of the entire universe. How could you *not* want to study it?

The Standard Model of Particle Physics is the best solution to the puzzle that has yet been postulated. A mind bogglingly accurate description of almost all known phenomena, it time and again comes up with the correct answer to every question we can throw at it. The Large Hadron Collider now smashes particles together at such high energy as to rewind 13.7 billion years and recreate the conditions of the universe fractions of a second after it's birth, and still the theory predicts the results of our searches and measurements to astonishing accuracy. However, the puzzle is not yet complete. The 2012 discovery of the Higgs boson [3, 4] filled one very important hole, but the Standard Model still has nothing to say on gravity, dark matter, or dark energy, while the issues of the fermion mass hierarchy, neutrino masses and the matter-antimatter asymmetry of the observable universe are not yet satisfactorily put to bed. We venture forth with little guidance on what it will look like; just our intuition, ingenuity, and unbridled belief that whatever it is, it is beautiful.

One way in which to approach the puzzle of what lies beyond the Standard Model is to make precise measurements of the fundamental parameters of the theory, such as the Yukawa couplings [5] between the fermions and Higgs boson. The Yukawa coupling for the top quark, y_t, is a particularly interesting parameter to measure: it is responsible for giving fermions mass, and the top quark is the most massive fundamental particle of all. Why should the top quark be so much more massive than the other fermions? Does it play a special role in electroweak symmetry breaking, or in whatever new physics is out there?

The only way to measure y_t directly is via associated production of top quarks and Higgs

bosons. However, this is an incredibly rare process, with top quark pair plus Higgs boson ($t\bar{t}H$) production being just 1% of the total Higgs boson production cross-section, and single top quark plus Higgs boson (tH) significantly smaller still [6]. The most abundant decay channel, the Higgs decay to a pair of b quarks, is buried under a huge background of top quark pair production, the cross-section of which is several orders of magnitude larger [7]. The problem then is two-fold: ensuring that the top quark is well understood, such that we better understand the role it might play in new physics, and may begin to hunt for the associated Higgs production needle amongst the top quark hay; and then designing a sufficiently efficient strategy such that we have a chance of finding that needle. It is these problems that this thesis seeks to solve, first via a measurement of the kinematics of top quark production, and subsequently via searches for the $t\bar{t}H$ process.

At the Large Hadron Collider (LHC), located on the outskirts of Geneva, Switzerland, proton-proton collisions took place between 2015 and 2018 at an unprecedented centre-of-mass energy, commonly denoted as $\sqrt{s}$, of 13 TeV. The ATLAS detector recorded data throughout this period, ending the second run of the LHC with 140 fb^{-1} of data with which to probe the Standard Model and search for clues on what lies just out of view.

This document is structured as follows. Chapter 2 gives a brief overview of the Standard Model of Particle Physics, paying particular attention to the aspects of the theory most relevant to the analyses presented. Chapter 3 introduces some of the more advanced concepts utilised in the thesis, including boosted particles, detector unfolding, profile likelihood fits, and multivariate analysis methods. The ATLAS detector, and LHC accelerator complex, are discussed in Chapter 4, and the particle physics objects reconstructed in the detector are described in Chapter 5. Monte-Carlo simulations, used extensively in the presented analyses, are introduced in Chapter 6. The first of the measurements presented, the differential cross-section measurement of boosted top quarks, is detailed in Chapter 7. Chapter 8 describes the first ATLAS $t\bar{t}H, H \to b\bar{b}$ paper to incorporate a boosted signal region, with Chapter 9 discussing the optimisations of the boosted region that have been developed since the initial publication. Finally, Chapter 10 summarises the results, and briefly looks to the future of field.

Chapter 2

The Standard Model of Particle Physics

> *There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.*
>
> – Douglas Adams

The Standard Model of Particle Physics (SM) is an astonishing theory. It is complex enough to describe phenomena both over a wide range of orders of magnitude and centre-of-mass energies, as shown in Figure 2.1. Indeed, the SM correctly predicts the electron magnetic moment to an incredible 14 decimal places [8]. Simultaneously, it is simple enough to be written on the side of a coffee cup, as in Figure 2.2.

The full SM Lagrangian can be written as the sum of four smaller pieces

$$\mathcal{L}_{\text{SM}} = \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + (|D_\mu\phi|^2 - V(\phi)) + (\psi_i y_{ij}\psi_j\phi + h.c.) \tag{2.1}$$

$$= \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Fermions}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}, \tag{2.2}$$

where $F_{\mu\nu}F^{\mu\nu}$ is the scalar product of the field strength tensor encapsulating the strong nuclear force and the combined electroweak force. $\bar{\psi}\not{D}\psi$ represents the dynamics of the fermion fields ψ, with $\not{D}$ the covariant derivative on fermions required to maintain gauge symmetry. D_μ is the covariant derivative on the boson fields, and thus $D_\mu\phi$ encodes the interaction of the gauge bosons with the Higgs field ϕ. $V(\phi)$ is the potential of the Higgs field, from which spontaneous symmetry breaking emerges. $\psi_i y_{ij}\psi_j\phi$ describes the interaction between the fermions ψ and the Higgs field, with y_{ij} the elements of the Yukawa matrix that encodes the strength of these interactions. Finally, $h.c.$ represents the Hermitian conjugate of the Yukawa terms, which are not self-adjoint. In essence, these terms respectively encapsulate: the description of the fundamental forces themselves; the dynamics of the fermions; the Higgs field and it's coupling to the gauge bosons; and the coupling between the Higgs field and the fermions.

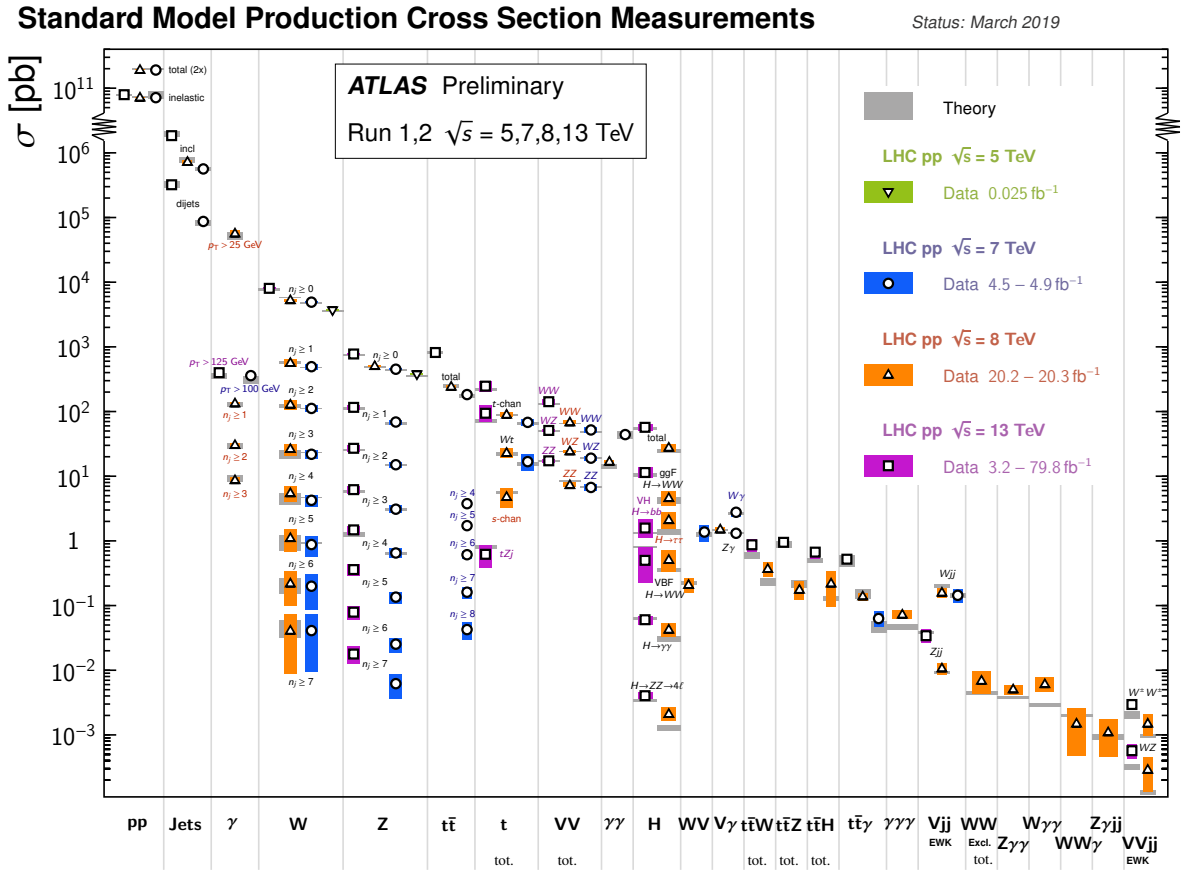


Figure 2.1: Summary plot of SM cross-section measurements, over many orders of magnitude and at varying centre-of-mass energies, as measured by ATLAS [9].

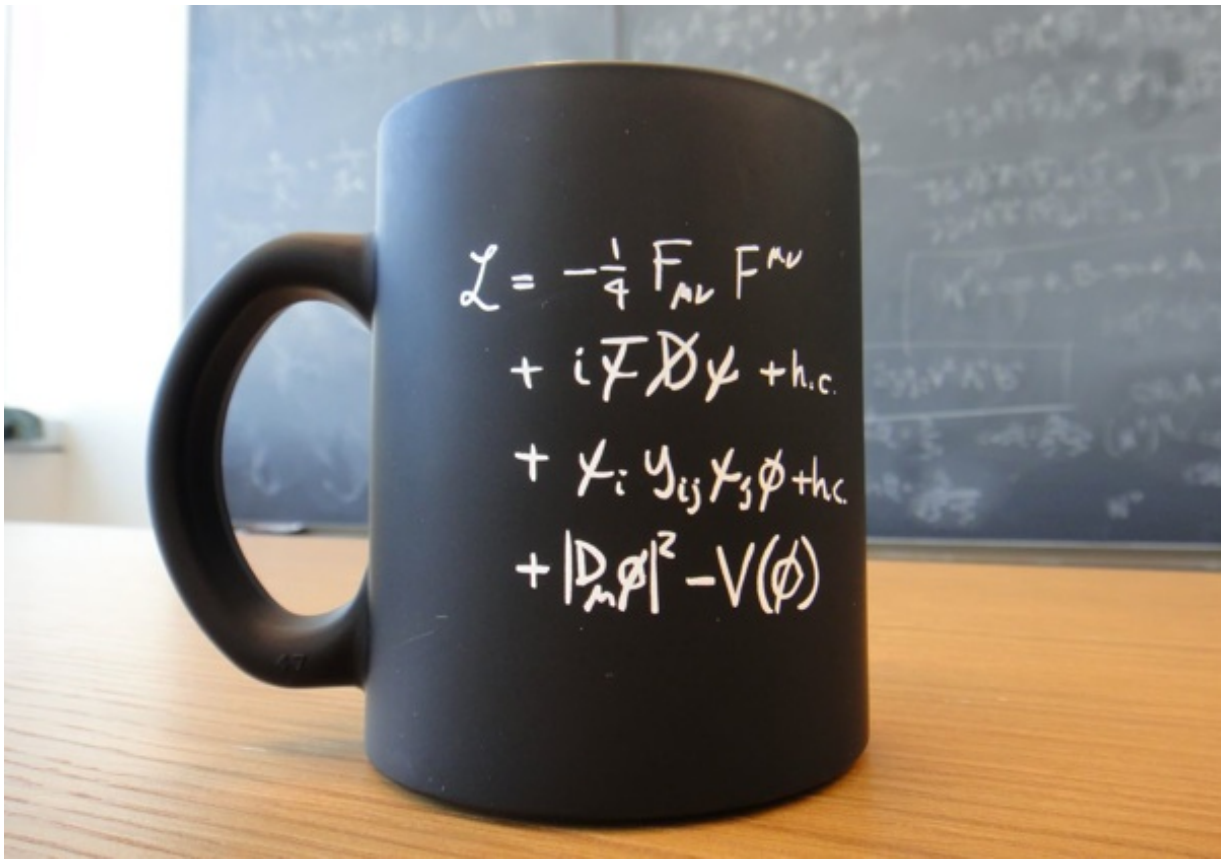


Figure 2.2: The SM Lagrangrian written on the side of a coffee cup, modulo one controversial mistake [10]. Available at the CERN gift shop!

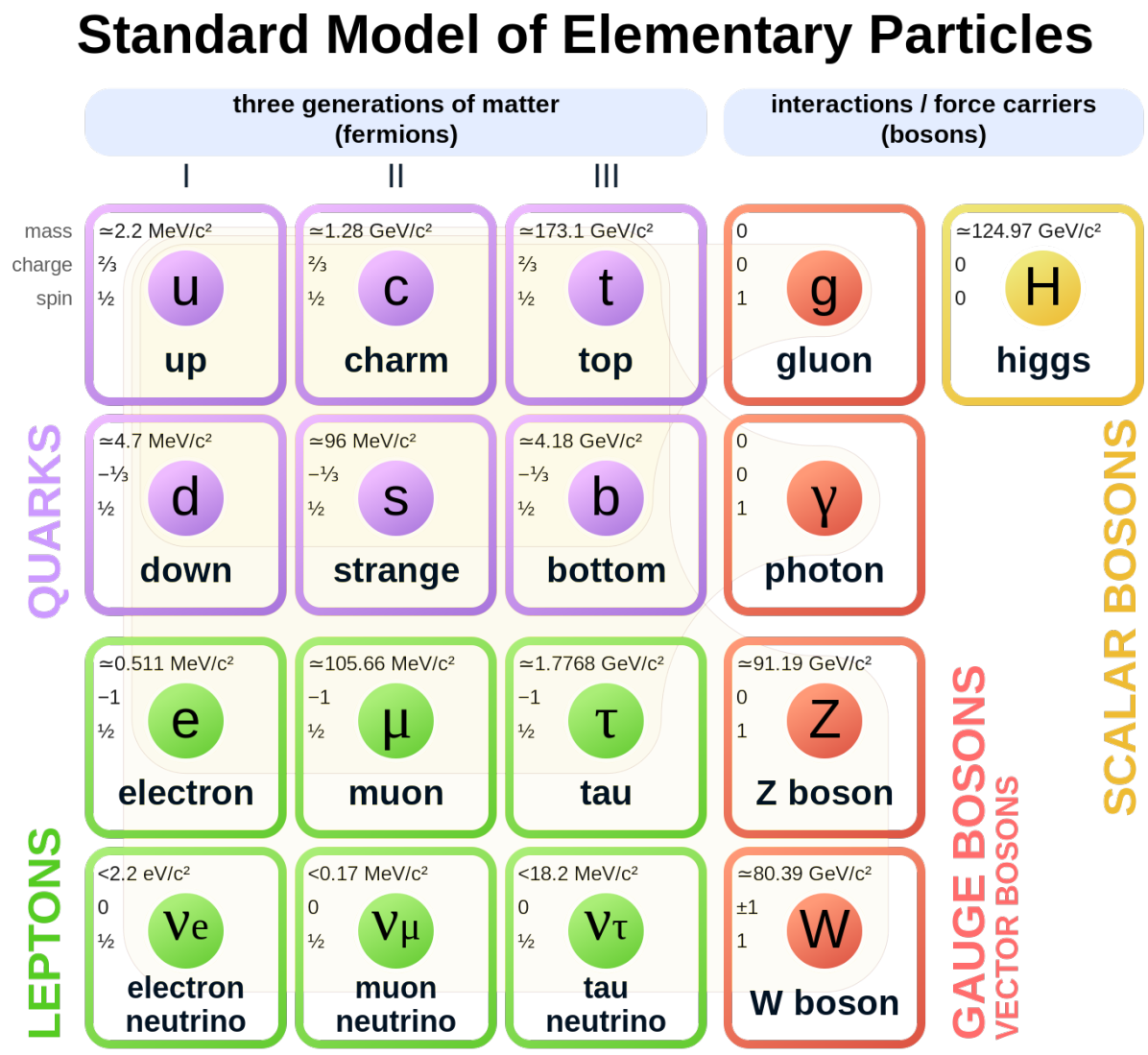


Figure 2.3: The particle content of the Standard Model of Particle Physics [11].

The full particle content of the SM is summarised in Figure 2.3. The following sections look at each of these smaller pieces in turn.

2.1 $\mathcal{L}_{\text{Gauge}}$: The Fundamental Forces of Nature

$$\mathcal{L}_{\text{Gauge}} = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{2.3}$$

The first piece of the SM Lagrangian, $\mathcal{L}_{\text{Gauge}}$, concerns the fundamental forces and the particles which carry them, the gauge bosons, as well as their interactions. The SM is constructed as a quantum field theory, in which the symmetries of nature are manifest as conserved currents, following Noether's theorem [12]. The SM thus relies on the gauge symmetry of the group

$G_{SM} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where the three conserved currents are the *color* C, the left-handed weak isospin L, and the weak hypercharge Y.

$SU(3)_C$ represents the strong interaction, and gives rise to *quantum chromodynamics* (QCD) [13, 14], the force governing the interactions of quarks and gluons. The three "colors" of quark, rgb, interact via gluons, which are the gauge bosons resulting from the eight generators of the group. The strength of the strong interaction is governed by the strong coupling constant α_s, which is in fact energy dependent and thus not a constant at all [15]. At LHC energies, α_S is sufficiently small as to be calculable perturbatively: see Chapter 6. The gluons themselves are color charged, electrically neutral, and massless. Being color charged means that gluons have a self-interaction, and this leads to the principle of quark confinement, whereby colored particles will always bind to form colorless states: at low energy, the self-interaction creates an anti-screening effect that causes the strength of the strong interaction to increase. Thus, any particles that have a net color charge are bound together to form color neutral states.

$SU(2)_L \times U(1)_Y$ describes *quantum electrodynamics* (QED), the unified theory of electrodynamics and the weak interaction [16–18]. $SU(2)_L$ and $U(1)_T$ have three and one *generator(s)*[1] respectively. The gauge symmetry nominally requires that the corresponding bosons are massless, but observation of massive gauge bosons $W^{\pm}$ and Z^0 implies something is missing: this is the third piece of the Lagrangian, $\mathcal{L}_{\text{Higgs}}$, discussed in Section 2.3. The mass of these particles causes the weak interaction to have a very short range. The fourth boson is the humble photon, which is responsible for the electromagnetic interaction. Being massless and electrically neutral the photon does not have a self-interaction and has an infinite range.

2.2 $\mathcal{L}_{\text{Fermions}}$: Quarks and Leptons

$$\mathcal{L}_{\text{Fermions}} = i\bar{\psi}\not{D}\psi \tag{2.4}$$

The fundamental building blocks of everything in the universe, from stars and galaxies to the paper or screen upon which you are reading this document, and indeed including the human body itself, are fermions: quarks and leptons. The protons and neutrons in the atomic nucleus are made of up and down quarks (bound together by gluons), with orbiting electrons completing the classical picture. The second piece of the SM Lagrangian, $\mathcal{L}_{\text{Fermions}}$, describes the dynamics of these particles. The covariant derivative $\not{D}$ encodes the interaction between the fermions, via the gauge fields which propagate the strong, weak, and electromagnetic forces.

In fact, there are 3 generations of fermions, as shown in Figure 2.3. The three generations of fermions have identical properties in terms of their color, weak, and electromagnetic charges, varying only in their mass. Quarks carry color charge, and therefore experience the strong force, while leptons do not. All fermions interact via the weak force, and all but the electrically neutral

[1] The generators of a group are the set of objects from which the full dynamics of the group emerge; in this case, they correspond to the set of gauge bosons which propagate the field.

neutrinos interact electrodynamically. The heavy fermions are unstable, and thus decay, with lifetimes inversely proportional to their mass. The heaviest fermion of all is the top quark.

2.2.1 The Top Quark

The top quark was first predicted by Kobayashi and Maskawa [19] in 1973 as part of an as yet unseen 3rd generation of fermions. Discovery of the bottom quark [20] and the τ lepton [21] followed in the same decade, yet the top quark remained absent. It was finally discovered in 1995 at the Tevatron based CDF [22] and D0 [23] collaborations at a centre-of-mass energy of 1.8 TeV. Due to its high mass, 172.69 GeV ± 0.25 (stat) ± 0.41 (syst) in the most recent ATLAS combination [24], it could not be produced by previous colliders. This unusually high mass leads to some interesting properties. With a lifetime of $\tau \sim 10^{-25}$ s [25] – shorter than the time required for hadronisation $\tau \sim 10^{-22}$ s [25] – it decays before it can form bound states. The top can thus be reconstructed and studied in experiments as a bare quark, something that is not possible for any other quark due to the effects of QCD confinement – a unique window through which to study the Standard Model.

At the LHC, top quark pairs are produced by the strong interaction: 90% of $t\bar{t}$ pairs are via gluon fusion, with the remaining 10% via $q\bar{q}$ annihilation, at $\sqrt{s} = 13$ TeV [26]. Tops decay almost exclusively into a W boson and a b quark [27]. The leading order Feynman diagrams for top quark pair production in a pp collider are shown in Figure 2.4. Single top quark production is also possible, though the cross-section is comparatively small due to being a weakly initiated process. The b quark binds with other quarks and produces a *jet*, a narrow spray of detectable hadronic particles. Since the W can decay either hadronically into a pair of quarks or leptonically into a charged lepton+neutrino, the $t\bar{t}$ events can be categorised into three channels based upon the W decays: dilepton, lepton+jets, or alljets, as shown in Figure 2.5. The dilepton channel has the smallest branching ratio of the three at about 11%, meaning much lower statistics are available. Though the leptons, which can be identified relatively easily, mean a good selection efficiency, the second neutrino in the final state makes the accurate reconstruction of the full event challenging. The branching ratio of the alljets channel is much larger, at around 45%, but suffers from a large background due to QCD multijet production. It is further harder to trigger on due to the lack of final state lepton, and the high jet multiplicity makes it a combinatoric challenge to correctly reconstruct the top quarks. Lepton+jets events have a branching ratio similar to the alljets, but include the key lepton for easy triggering and lower backgrounds, as well as a smaller combinatorics problem.

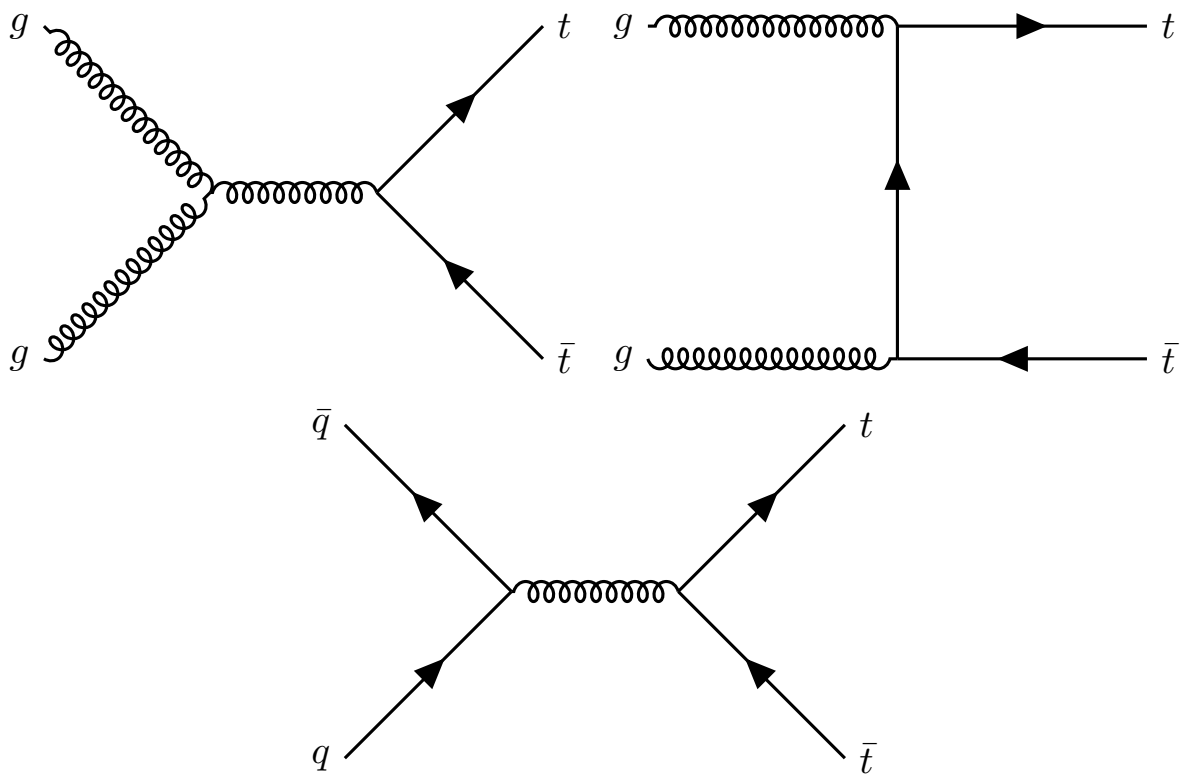


Figure 2.4: Leading order Feynman diagrams for $t\bar{t}$ production at the LHC.

2.3 $\mathcal{L}_{\text{Higgs}}$: Electroweak Symmetry Breaking and the Higgs Mechanism

$$\mathcal{L}_{\text{Higgs}} = |D_\mu \phi|^2 - V(\phi) \tag{2.5}$$

In order to generate mass terms for the gauge bosons of the weak force – without breaking gauge symmetry – an additional scalar field is required. The Brout-Englert-Higgs mechanism, first postulated in 1964 by Higgs [29] [30], Englert and Brout [31], and Guralnik, Hagen and Kibble [32], does exactly this, introducing a new scalar field ϕ and potential of the form $V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$, with μ^2 and λ free parameters. If $\mu^2 < 0$ and $\lambda > 0$, the potential is at its minimum at $V(\phi) = -\mu^2/2\lambda \equiv \upsilon$. This value υ is known as the *vacuum expectation value* of the field. There are a ring of possible ϕ values, as shown in Figure 2.6, that reach this minimum, and when the field is displaced from zero, the (arbitrary) choice of which value in this ring is taken leads to the description of the symmetry as being *spontaneously* broken. Mass terms corresponding to the $W^\pm$ and Z^0 boson now appear, along with an additional degree of freedom with units of mass – an excitation of the scalar field, the Higgs boson (H), with mass $m_H = \sqrt{-2\mu^2}$. The observed masses of the $W^\pm$, Z, and H bosons lead to a vacuum expectation value of $\upsilon \approx 246$ GeV. λ corresponds to the strength of the Higgs boson self-coupling, which is as yet unobserved [33].

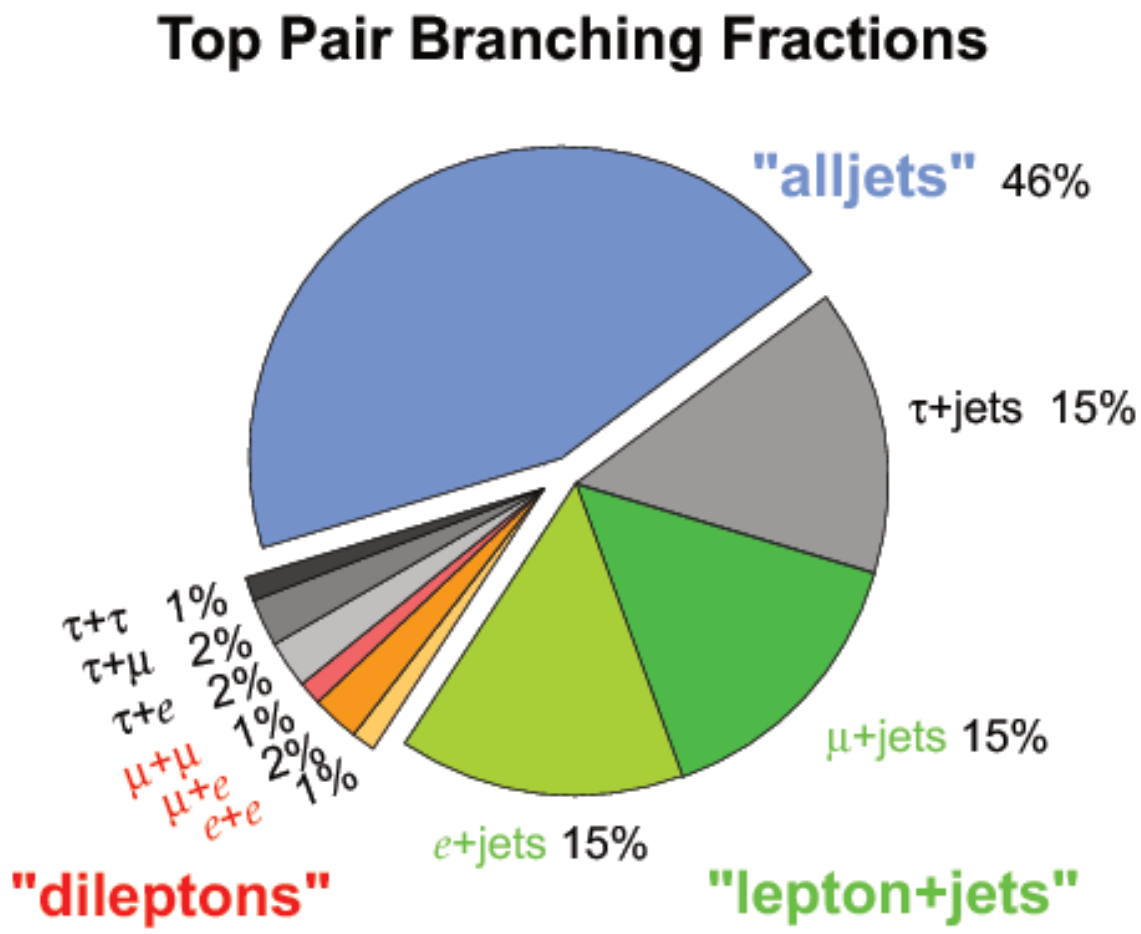


Figure 2.5: Breakdown of top quark final states [28]. The analyses presented in this thesis focus mainly on the e+jets and μ+jets decays, together referred to as the "lepton+jets" channel.

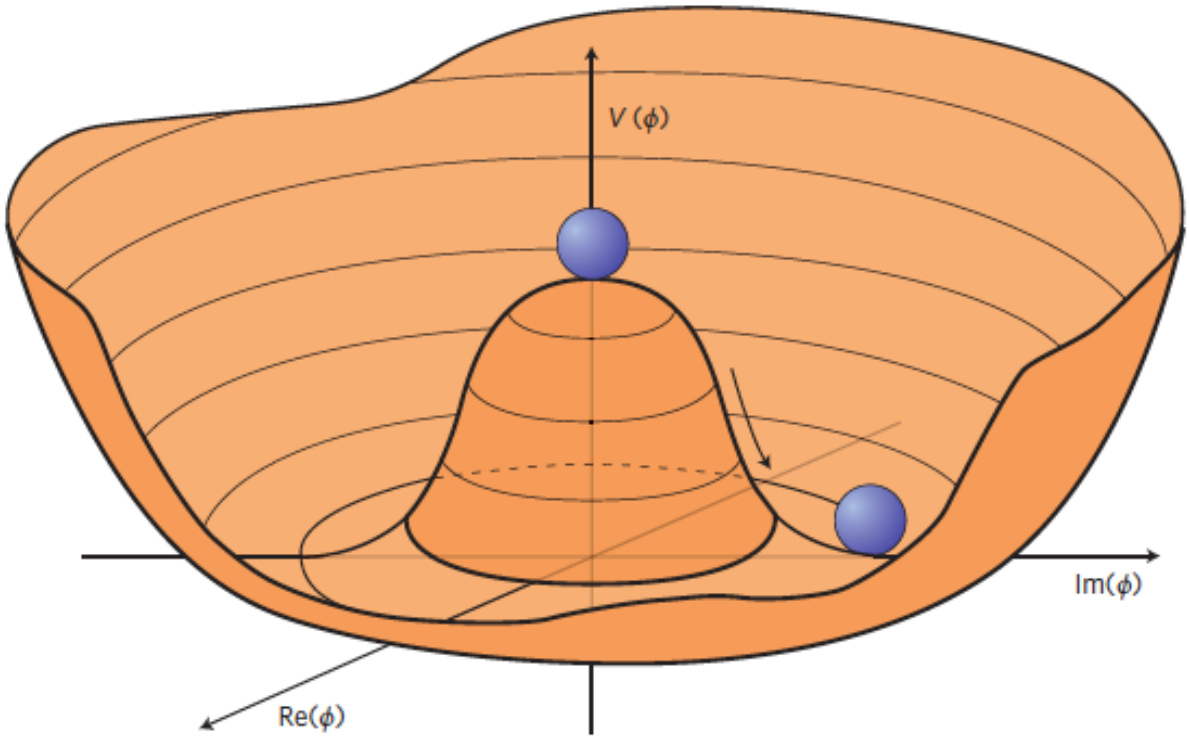


Figure 2.6: Graphical representation of the Higgs potential. The motion of the ball falling into the minima of the potential, and the resulting value in the plane orthogonal to the $V(\phi)$ axis, represents the "spontaneous" symmetry breaking [34].

2.3.1 The Higgs Boson

The mediating particle of the new scalar field, the Higgs boson, was discovered in 2012 by both ATLAS [3] and CMS [4]. Since then, precision measurements of the particle have found it to be consistent with the SM, with a mass $m_H = 125.09 \pm 0.24$ GeV as of the latest ATLAS/CMS combination [35]. Higgs bosons are produced in four primary production modes at the LHC: gluon-gluon fusion (ggF), vector boson fusion (VBF), associated production with a vector boson (VH), and associated production with a pair of heavy quarks ($t\bar{t}H$and $b\bar{b}H$), Feynman diagrams for which are shown in Figure 2.7. The cross-sections of each of these processes as a function of centre-of-mass energy are shown in Figure 2.8.

The Higgs boson couplings to other particles are directly related to the mass of those particles. The branching ratios of the Higgs is thus highly sensitive to the Higgs boson mass, as shown in Figure 2.9a. At the observed value $m_H \approx 125$ GeV, shown in Figure 2.9b, the dominant decay mode is to a pair of bottom quarks, with a total of 57% of all decays in this channel. WW and ZZ decays are possible only if one of the two vector bosons are off shell, suppressing these channels. Higgs decays to two gluons or photons are also possible despite these particles being massless, due to loop diagrams (similar to the ggF production diagram in Figure 2.7a), with the largest contribution from the top loop. Since the top quark is more massive than the Higgs boson itself,

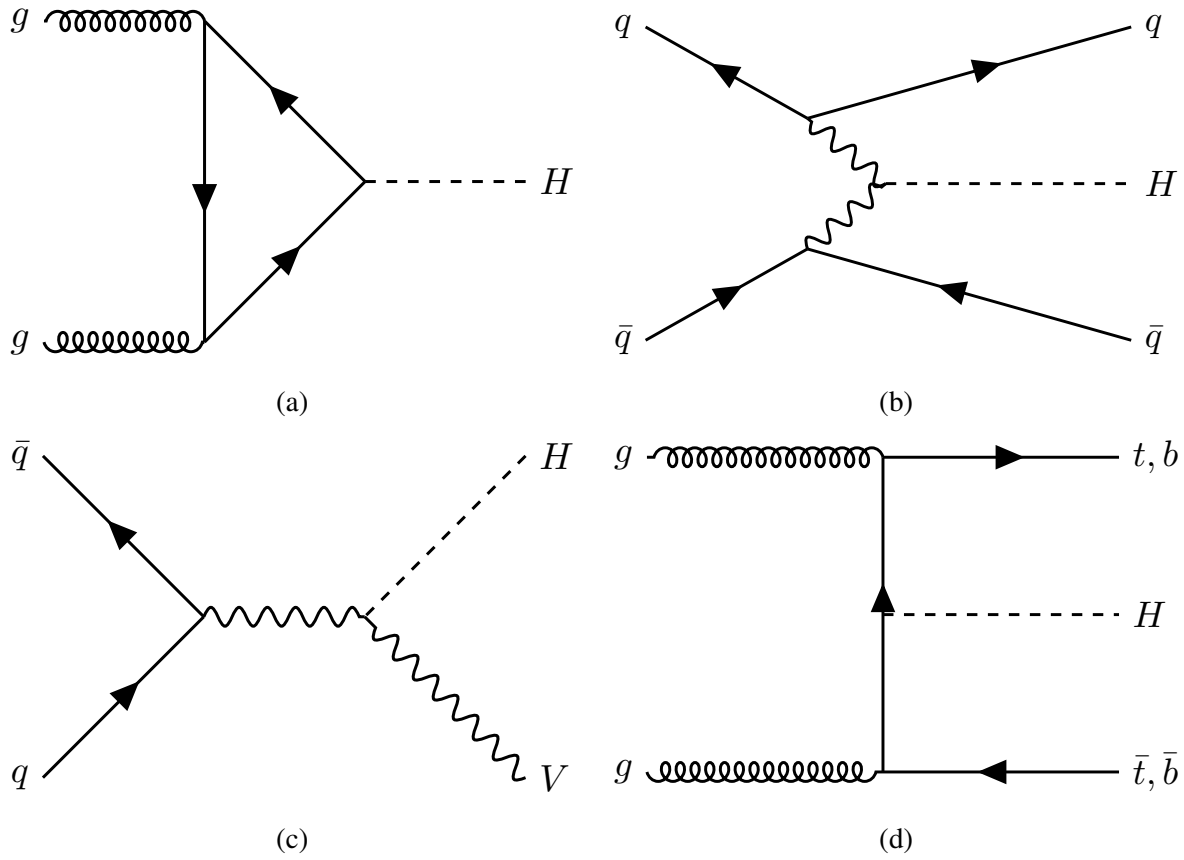


Figure 2.7: Representative Feynman diagrams for the main Higgs boson production modes at the LHC, (a) gluon-gluon fusion, (b) vector boson fusion, (c) associated production with a vector boson, (d) associated production with a pair of top or bottom quarks.

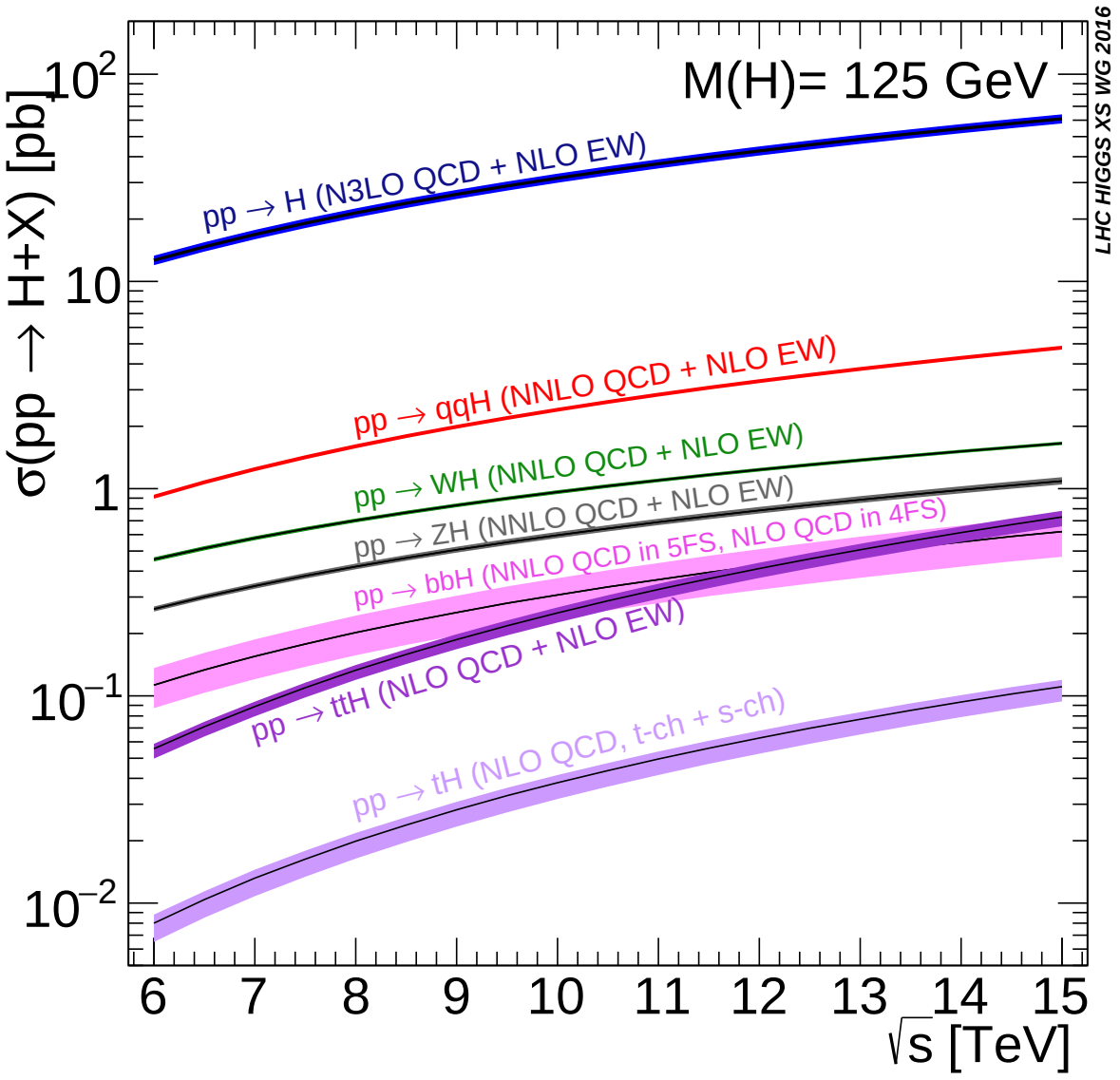
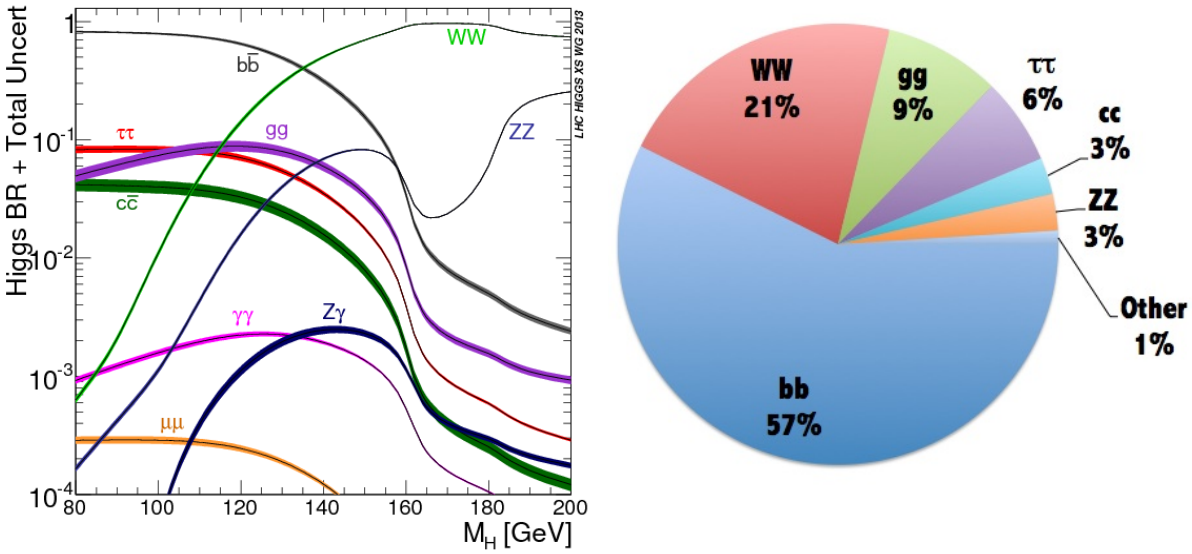


Figure 2.8: Higgs boson production cross-sections in the SM, as a function of centre-of-mass energy [6].



(a) (b)

Figure 2.9: Branching ratios for the SM Higgs boson, (a) as a function of m_H [36], and (b) at the observed value of $m_H = 125$ GeV [37]. The branching ratios are highly mass dependent, with the dominant decay at $m_H = 125$ GeV being $H \to b\bar{b}$.

the Higgs boson decay to a pair of top quarks is kinematically forbidden.

2.4 $\mathcal{L}_{\text{Yukawa}}$: Acquiring Fermion Masses

$$\mathcal{L}_{\text{Yukawa}} = \psi_i y_{ij} \psi_j \phi + h.c. \tag{2.6}$$

The new scalar field also introduces the terms shown in Equation 2.6, which are responsible for the fermion masses. In the SM, the fermion masses m_f are proportional to υ via the relation

$$m_f = y_f \frac{\upsilon}{\sqrt{2}}, \tag{2.7}$$

where y_f is the size of the coupling between the Higgs boson and the fermion f, known as the *Yukawa coupling*. The size of these couplings are not predicted by the SM, but are free parameters of the theory. This relation demonstrates that the Higgs couples more strongly to more massive particles, or equivalently that particles which couple more strongly to the Higgs have a larger mass. Inserting the observed top quark mass and υ into Equation 2.7 gives $y_t \approx 1$. This result has lead to speculation that the top quark plays a special role in electroweak symmetry breaking and physics beyond the Standard Model, which is discussed in Section 2.5. To date, only indirect measurements of y_t have been possible; measurements of ggF production and $H \to \gamma\gamma$ decays are sensitive to y_t via the top quark contribution to the loop diagrams, though relies on the assumption that there are no BSM particles entering these loops. The combined ATLAS + CMS best fit for y_t in this way is $1.40^{+0.24}_{-0.21}$ times the SM expectation [38]. The CMS Collaboration

also recently published an indirect measurement of y_t, via differential distributions of top quark kinematics [39], though this measurement is less sensitive than the ggF and $H \to \gamma\gamma$ analyses with a best fit of $1.07^{+0.34}_{-0.43}$ times the SM. A direct measurement of the top-Higgs coupling y_t is thus of extremely high interest as a test of the SM.

2.4.1 The Top-Higgs Coupling

Since the Higgs decay to two top quarks is heavily suppressed, the top Yukawa can only be measured directly via associated production of the two particles. A representative leading order Feynman diagram for the associated production of a Higgs boson with a pair of top quarks, $t\bar{t}H$, is shown in Figure 2.7d. Associated production with a single top quark, tH, is heavily suppressed in the SM by the interference between diagrams [40], and so is not discussed in detail here.

In spite of this special interest and a huge effort, both ATLAS [41–43] and CMS [44,45] were unable to detect any significant excess of $t\bar{t}H$ production in Run 1, due to the low production cross-section of such a high mass event and the irreducible background from $t\bar{t}$ events. At higher centre-of-mass energies, the SM cross-section of the $t\bar{t}H$ production processes increases faster than the backgrounds: by a factor of 7.1, from 86 fb at 7 TeV to 611 fb at 14 TeV [46], whilst the cross-section of the main $t\bar{t}$+jets background increases from 172 pb at 7 TeV to 953 pb at 14 TeV [7]. This means the signal to background ratio will be improved in Run 2, with data taking at 13 TeV. Additionally, with the higher energy available, a higher fraction of these events will be highly energetic, motivating a search in the high p_T *boosted* regime that is described in Section 3.1. The boosted channel is also potentially more sensitive to deviations from the SM at the LHC [47], due to this higher energy.

2.5 $\mathcal{L}_{\mathrm{BSM}}$: Beyond the Standard Model

Despite it's success, the Standard Model is an incomplete theory, being unable to describe many observed phenomena. The astute reader will have noted the absence of any mention of the force of gravity from Section 2.1, and indeed the hunt for a quantum theory of gravity that could incorporate it into the SM has so far evaded us. There is no prior reason why there should be three generations of fermions, or why the masses of these generations are at such different scales: why is $m_t >> m_u$? It similarly has nothing to say on the topic of dark matter or dark energy, nor does it contain enough charge-parity violation to account for the huge matter-antimatter asymmetry observed in the universe. The observation of neutrino oscillations implies a non-zero mass for these particles, implying the existence of as-yet undetected right-handed neutrinos, or some other new physics [27].

At one loop, the Higgs boson should pick up large quantum corrections to its mass, dominated by the top quark loops. With Λ the energy scale at which new physics becomes apparent, this

correction is approximately [48]

$$\delta m_H^t \sim \frac{-m_t^2}{v^2}\frac{\Lambda}{4\pi^2}. \tag{2.8}$$

Thus, the larger the scale of new physics, the larger the Higgs boson mass correction is. With Run 2 of the LHC already probing the TeV scale, no new physics observed, and the observation of the Higgs mass at $\sim$ 125 GeV, a large amount of *fine-tuning* is required to cancel out these contributions to match observation. This is often referred to as the *hierarchy problem* or *naturalness* argument. Many new physics scenarios predict top-partners which lead to cancellation of these contributions, such as those of supersymmetry [49, 50] or vector-like quarks [51]. With no sign of TeV scale new physics, and with no large energy increase on the particle physics horizon, it is increasingly unlikely that any new particles will be observed in the near future. It is thus incumbent on the discerning physicist to probe for further holes in the SM, which might provide clues as to the nature of physics beyond the SM (BSM). This can be achieved by increasingly precise measurements, which may deviate slightly from the SM predictions. As the heaviest particle in the theory, the top quark may reasonably be expected to play an important role in physics at high energy scales, such as through anomalous couplings in the top sector [52, 53] or higher order operator effects [54, 55], and the effects of this may be visible via deviations from the SM at LHC energies. All of these possible BSM physics scenarios motivate the measurement of top production presented in Chapter 7, as well as the $t\bar{t}H$ analyses in Chapters 8 and 9. The measurement of the top Yukawa via $t\bar{t}H$ is further motivated by composite Higgs models [56], anomalous Higgs couplings [57, 58], and has implications for the stability of the electroweak vacuum [59].

Chapter 3

Analysis Techniques

Solutions nearly always come from the direction you least expect, which means there's no point trying to look in that direction because it won't be coming from there.

– Douglas Adams

This chapter introduces a number of common analysis techniques that are used in the analyses presented in Chapters 7, 8, and 9. First, the concept of *boosted particles*, which are used throughout, is introduced in Section 3.1. Section 3.2 describes the method of *unfolding*, whereby detector effects are removed from the measured data, important for the results in Chapter 7. *Profile likelihood fits*, used extensively in both Chapter 8 and Chapter 9, are then described in Section 3.3, followed in Section 3.4 by introductions to the *multivariate analysis* techniques used in a number of places in the same chapters.

3.1 Boosted Particles

With the increase in centre of mass energy in Run 2 of the LHC, more particles than ever before are produced at high transverse momentum, including high mass particles such as top quarks and Higgs bosons. Traditionally, the decay products of these particles are reconstructed individually: top quarks decaying $t \to Wb \to qqb$ and Higgs bosons decaying $H \to b\bar{b}$ would typically be reconstructed as three or two approximately conic sprays of mostly hadronic particles, known as *jets* – discussed further in Section 5.3. This is referred to as the *resolved* topology. At high p_T, these decay products become collimated, eventually causing the jets to overlap and become unresolvable. An alternate topology, referred to as the *boosted* regime, utilises this collimation by reconstructing high p_T heavy particles as a single larger radius jet, as demonstrated in Figure 3.1. Given the rule of thumb $\Delta \approx \frac{2m}{p_T}$ for the angular separation between the decay products of a particle of mass m, a jet of radius $R = 1.0$ will typically fully contain a top quark with $p_T > 350$ GeV or a Higgs boson with $p_T > 250$ GeV. The advantage of this technique is that it removes the combinatoric ambiguity of which jet originated from which decay product, while

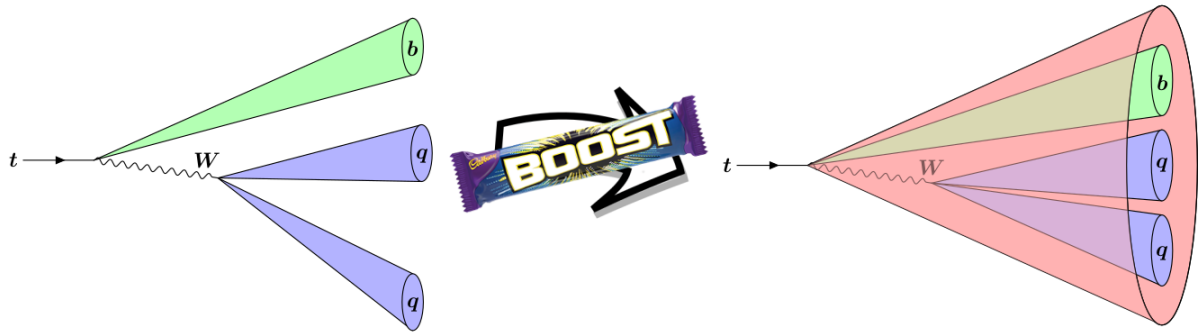


Figure 3.1: Cartoon depiction of the transition between the resolved and boosted regimes, for the case of a hadronically decaying top quark.

jet substructure techniques (see Section 5.3.5) can be utilised to tag the jet and further reduce background. In addition, the selection efficiency increases relative to the resolved regime as p_T increases, and thus the energy reach of measurements and searches increases. More on the jet tagging algorithms used in this thesis can be found in Section 5.3.6 and Section 9.3.

3.2 Unfolding

In ATLAS, Monte-Carlo (MC) simulations are passed through a simulation of the ATLAS detector, as described in Section 6.7, in order to compare data and MC at *detector level*. In order for the data to be useful to someone outside the ATLAS collaboration, it is important to remove effects due to the interaction with the detector hardware. The process of doing this is referred to in particle physics as *unfolding*, though in other applications like imaging or data science is often referred to as "deconvolution" or "unsmearing". Without unfolding, data from different experiments could not be compared directly to each other, and theorists would be required to pass their latest predictions through similar complex detector simulations – making unfolding a crucial analysis technique. However, the procedure of removing the detector effects from the data is not straightforward. Starting with the data measurement described by the function $g(d)$, the *truth level* distribution $f(t)$ is desired. The unfolding procedure is described by a Fredholm equation of the first type [60]

$$\int_{\Omega} K(d,t)f(t)dt + b(d) = g(d), \tag{3.1}$$

with $K(d,t)$ the kernel representing the detector effects, and $b(d)$ representing background events. The equation in this form can be thought of as representing the "folding" of the detector effects $K(d,t)$ into the truth distribution $f(t)$ to give the detector level distribution $g(d)$, analogous to the detector simulation step described in Section 6.7. We thus desire a methodology for inverting this operation. In this thesis we measure binned distributions $g(d)$ and $f(d)$, and in this case there are 5 basic types of convolutions to the data that can affect the measurement. These are:

(1) Migrations: when a measurement of the quantity is slightly smeared by limited detector resolution or similar, causing "migrating", where the detector level quantity falls into the wrong bin of the histogram.

(2) Efficiency: an event which passes all selection requirements at truth level, but fails one or more of the selection requirements at detector level.

(3) Acceptance: an event which passes all selection requirements at detector level, but fails one or more of the selection requirements at truth level.

(4) Background: the wrong type of event passing event selection, such as a W+jets event in a $t\bar{t}$ measurement.

(5) Statistical fluctuations: unavoidable fluctuations in the number of data events.

A number of methods for performing the unfolding exist, including simple bin-by-bin corrections [61], Fully Bayesian Unfolding [62], and Singular Value Decomposition [63]. In the analysis presented in Chapter 7, an Iterative D'Agostini [64] method, as implemented in the RooUnfold package [65], is used.

To remove effects like (4) above, the number of background events from simulations and data driven estimates of W+jets, Z+jets, Diboson, QCD Multijet, $t\bar{t}$+V and Single Top that pass selection are subtracted from each bin of the data. Next, an *acceptance* correction is applied, as calculated from the $t\bar{t}$ signal MC. This accounts for effects of type (3), where an event is selected at detector level but not particle level. These events are not desirable, as if they fail the particle level selection, they lie outside the fiducial phase space, and thus this factor should be ≤ 1. This correction factor is applied per bin, and is defined as the fraction of events which pass both the detector and truth level selections, divided by the events that pass detector level

$$f_{\text{acc}}^{j} = \frac{N_{\text{det\&truth}}^{j}}{N_{\text{det}}^{j}}, \tag{3.2}$$

where $N_{\text{det\&truth}}$ is the events which pass at both detector and truth level, N_{det} is the events which pass detector level, and the index j represents the bin in the detector level distribution.

Similarly, type (2) effects are corrected for using an efficiency correction ϵ, which accounts for events that passed at particle level but not at detector level, defined as

$$\epsilon^{i} = \frac{N_{\text{det\&truth}}^{i}}{N_{\text{truth}}^{i}}, \tag{3.3}$$

where N_{truth} is the events which pass at truth level, and in this case the index i is over the bins in the truth level distribution.

Finally, a *response matrix* $\mathcal{M}$ is built. This is a two dimensional distribution with the detector level on one axis and the truth distribution on the other, filled only with events that pass both

selections. Any events which fall off the diagonal of this distribution are thus of type (1) as described above. A *migration matrix* is a representation of the response matrix, where the bin contents are normalised along each row or column, to display the fraction of events which lie on or off diagonal.

The final equation governing the unfolding procedure is then given by

$$\frac{d\sigma^{\text{fid}}}{dX^i} = \frac{1}{\mathcal{L} \cdot \Delta X^i} \cdot \frac{1}{\epsilon^i} \cdot \sum_j \mathcal{M}_{ij}^{-1} \cdot f_{\text{acc}}^j \cdot (N_{\text{reco}}^j - N_{\text{bkg}}^j), \tag{3.4}$$

where σ^{fid} is the cross-section in the fiducial phase space, X is the quantity of interest, $\mathcal{L}$ is the integrated luminosity of the data, and ΔX^i is the width of bin i. The unfolding can thus be thought of as a matrix inversion problem[1], analogous to the inverse operation of the integration of the kernel $K(d,t)$ that embeds the detector response in Equation 3.1. To perform this matrix inversion, Bayes theorem can be used

$$P(T|D) = \frac{P(D|T) \times P(T)}{P(D)}, \tag{3.5}$$

where D represents detector level and T represents truth level. $P(D)$, the probability distribution for the data, here represents the "prior", the assumptions made on our knowledge of the problem. The resulting spectrum $P(T|D)$, the truth probability distribution given the data distribution, is referred to as the "posterior". $P(T)$ is the probability distribution of the truth, with $P(D|T)$ the probability distribution of the data given the truth. In the D'Agostini method, the initial chosen prior is the nominal MC distribution. In order to minimise the influence of this assumption, the matrix inversion is performed iteratively, with the posterior of the previous iteration used as the prior for the subsequent iteration. In this way, the bias of the SM assumption in the initial prior is minimised. The number of iterations performed is then a parameter of the unfolding that must be chosen – too many iterations will cause fluctuations caused by limited MC statistics to become increasingly influential in the unfolding, while too few will cause the SM prior to be influential and potentially mask new physics effects.

3.3 Profile Likelihood Fit

If, instead of unfolding measurements of a particular signal process in a signal dominated region, one wishes only to infer the presence of a signal over a potentially large background, a *profile likelihood fit* can be used. From each set of particle physics events $\boldsymbol{x}$, many parameters may be inferred. These parameters may represent a quantity to be measured, such as the cross-section of a particular process, referred to as the *parameter of interest* μ, or parameters, $\boldsymbol{\theta}$, the systematic

[1]This though is not strictly correct in D'Agostini unfolding, as it implies a single predetermined result could be analytically obtained. In truth, the problem is a probabilistic one: for further details, see [64]

uncertainties on the measurement of that quantity. These systematic uncertainties, or *nuisance parameters*, hinder the precision of the desired measurement. By considering N independent events $\boldsymbol{x} = x_1, x_2, ..., x_N$ each as a sampling from a probability density $f(x; \mu, \boldsymbol{\theta})$, an estimate of the values for μ and $\boldsymbol{\theta}$ can be built using the *likelihood function*

$$L(x; \boldsymbol{\theta}) = \prod_i^N f(x_i; \mu, \boldsymbol{\theta}). \tag{3.6}$$

The maximum value of this function occurs at $\hat{\mu}, \hat{\boldsymbol{\theta}}$, the best estimate for the parameters $\mu, \boldsymbol{\theta}$. Thus, to estimate μ and $\boldsymbol{\theta}$, the task is to maximise the likelihood function. In practice, it is often easier to minimise the negative log-likelihood $-\ln L$, which effectively replaces the product with a sum

$$-\ln L = -\sum_i^N \ln f(x_i; \mu, \boldsymbol{\theta}). \tag{3.7}$$

The minimisation of this function is not in general analytically calculable and so instead numerical techniques are used. In the fits described in this thesis, the MINOS method of the *minuit* program, as implemented in the RooFit package, is used [66].

In the case that there is a single parameter of interest μ and many nuisance parameters $\boldsymbol{\theta}$, the uncertainty on $\hat{\mu}$ can be found using a *profile likelihood ratio*

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}. \tag{3.8}$$

The double hat notation in $L = (\mu, \hat{\hat{\boldsymbol{\theta}}})$ denotes the maximum likelihood estimate of $\boldsymbol{\theta}$ at a given value of μ, while the single hat notation in $L = (\hat{\mu}, \hat{\boldsymbol{\theta}})$ denotes the *global* maximum likelihood estimate of μ and $\boldsymbol{\theta}$ together. With this, the compatibility between the data and any given value of μ can be easily tested, using the *test statistic* t_μ

$$t_\mu = -2 \ln \lambda(\mu). \tag{3.9}$$

The probability density function of the test statistic t_μ then follows a χ^2 distribution, and the uncertainty on $\hat{\mu}$ can be calculated using confidence intervals [67].

In practice, due to the large number of events under consideration in high energy physics, a profile likelihood fit is usually performed on a set of binned data. This data may be from many different orthogonal *regions*, defined by the set of fiducial requirements applied to the event. Some regions, referred to as *signal regions*, are selected to be enriched in signal events. Other regions, orthogonal to the signal regions and selected to target background processes, are referred to as *control regions*. Control regions allow a handle on the overall background rate and other nuisance parameters. The signal can then be measured as the amount of events

above the background expectation in the signal regions. It is common to further fit to a binned discriminant distribution in the signal regions that separates the background and signal processes, in order to maximise the sensitivity to the signal. In the case that the parameter of interest is the cross-section of a particular process, it is common to define $\mu = \sigma/\sigma_{\text{SM}}$, the ratio of the measured cross-section to the SM expectation. This acts as a normalisation factor of the signal event rate. Similarly, nuisance parameters $\boldsymbol{k}$ are assigned to the background rates, which may be constrained by the results of dedicated measurements of these processes or freely floated if a suitable such measurement does not exist. The expected total number of events $N_{\text{tot}}^{\text{exp}}$ in a given bin with n different background processes can thus be expressed as

$$N_{\text{tot}}^{\text{exp}}(\mu, \boldsymbol{k}, \boldsymbol{\theta}) = \mu \times N_{\text{sig}}^{\text{exp}}(\boldsymbol{\theta}) + \sum_{i}^{n} k_i \times N_{\text{bkg,i}}^{\text{exp}}(\boldsymbol{\theta}). \tag{3.10}$$

The nuisance parameters may in general follow any possible distribution. However, thanks to the central limit theorem, a Taylor expansion around the maximum point $\hat{\theta}$ may be assumed to be Gaussian in the high statistics limit [68]. The data distribution is assumed to be Poissonian. Thus, the full expression for the likelihood, divided into M bins with N_m^{obs} events in that bin, is

$$L(\mu, \boldsymbol{k}, \boldsymbol{\theta}) = \prod_{m}^{M} \frac{N_{\text{tot}}^{\text{exp}}(\mu, \boldsymbol{k}, \boldsymbol{\theta})^{N_m^{\text{obs}}}}{N_m^{\text{obs}}!} \times e^{N^{\text{exp}}(\mu, \boldsymbol{k}, \boldsymbol{\theta})} \times \prod_{i} \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta_i^2}{2}}. \tag{3.11}$$

The choice of region definitions, discriminant distributions, and the set of nuisance parameters, are referred to as the *fit model*. The fit model is usually optimised on *Asimov data*, in order to not be biased by fluctuations in the real data. An Asimov dataset is an artificial one in which the number of events in each bin in the fit is exactly equal to the expected number of events, and as such the number of events may be non-integer [67]. This is achieved using MC estimates for signal on top of the data-driven and MC estimates of all of the backgrounds.

3.4 Multivariate Analysis

In many cases, simple linear requirements on observables are not sufficient to obtain clear separation between two processes – for example to isolate $t\bar{t}H$ production from $t\bar{t}$ backgrounds, or for selecting jets initiated from b-quarks against the background of c-quark, light-quark and gluon initiated jets. In recent years, it has become increasingly common in high energy physics to utilise *machine learning* (ML) techniques. In particular, *multivariate analysis* (*MVA*) machine learning techniques have seen very widespread use. In these cases, many different observables may provide some small amount of separation between each case, and MVAs may allow the automated determination of the optimal set of requirements, or exploitation of non-linear correlations between different observables in order to identify the most likely origin of a pp

event, jet source, or any other such classification problem[2]. Currently, two of the most popular of these MVA algorithms are *Boosted Decision Trees* (*BDT*) [69] and *Deep Neural Networks* (*DNN*) [70].

3.4.1 Common Concepts

There are many concepts common to the MVA algorithms described here. In high energy physics, it is common to use simulated events to train these algorithms, since a true class label can then be assigned to every event. This is known as *supervised learning*, though some examples of *weakly supervised* or *unsupervised* learning, where the truth labels are not known, are also becoming relevant [71,72]. The MVA is given some set of *input features*, which are properties of the object to be classified, and the output of the MVA then discriminates between the different classes used in the training.

Any algorithm which requires training is at some risk of *overtraining* or *overfitting*, whereby the performance on the data used to train the algorithm exceeds the performance on an independent dataset. In this case, it is clear that the algorithm has learned features that are present in the training set which do not generalise to other datasets. This could be due to finite training statistics, or a non-representative sample. The former is a common problem in high energy physics, due to the high CPU overheads on many aspects of event simulation. The latter might also happen in the case that the simulation is not representative of the data, and thus only features which are well modelled should be utilised, and careful checks must be carried out to ensure performance is reliable and understood.

Evaluating the performance of any MVA should always be performed on an independent dataset which was not used in the training, to avoid overly optimistic evaluations due to overtraining. It is usual to partition the given dataset into *training* and *testing* datasets. Since the MC events used to train are usually required to be used in the analysis itself too, it is common to use *cross-training*, whereby the data is split into two equal sized subsets and two separate MVAs are trained with each half and tested with the other. Comparing the performance of these two independent MVAs is then an important check of overtraining, and the real data will simply use one of the two, chosen at random event by event.

Overtraining can often be mitigated by careful choice of *hyperparameters*. Hyperparameters are parameters of the algorithms themselves, which affect the learning of the algorithm. They must be chosen by hand, or otherwise optimised in some fashion, and what works best varies is highly problem dependent. The nature of the hyperparameters can take a wide variety of forms, and some common examples are discussed in detail for the case of BDTs and DNNs in Sections 3.4.2 and 3.4.3.

[2]One may also consider MVA regression, though this is not performed in any of the presented analyses and so not described here.

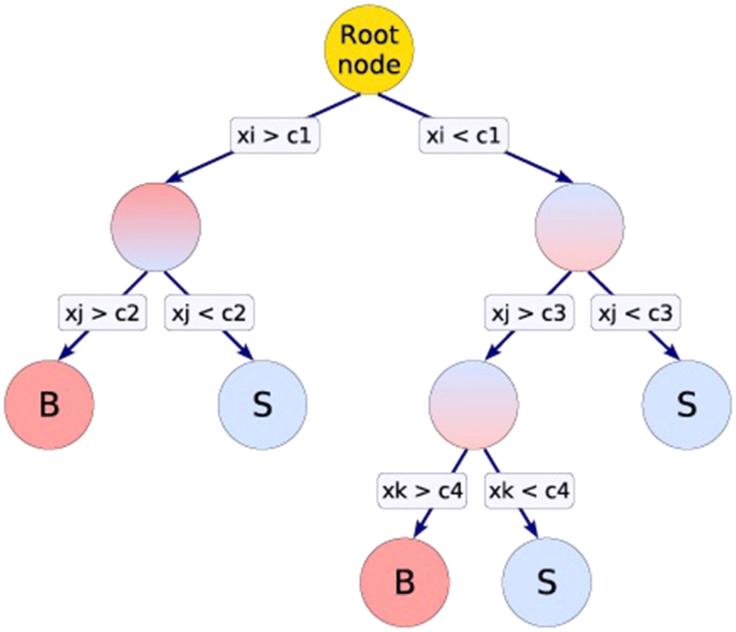


Figure 3.2: Example schematic of a Decision Tree [74]. Each circle represents a node, with the labels S and B representing signal dominated and background dominated nodes. The notation $c1$, $c2$, $c3$, and $c4$ represent the set of splitting thresholds applied at each branching, with xi, xj, and xk the input quantities that are being selected on.

3.4.2 Boosted Decision Trees

Boosted Decision Trees are an extension of the concept of a *decision tree* [73]. Given a set of input features with labelled classification, a decision tree will find the location of the optimal threshold on one of these observables in order to separate the classes. Such a splitting is referred to as a *node*, with the subsequent paths referred to as *branches*. Most commonly, BDTs are used as binary classifiers, though it is also possible to have more than two possible labels. The process may continue indefinitely, splitting the daughter nodes further until full separation is obtained, or else complexity can be limited by terminating the splitting at some maximum depth or minimum node size. The resulting end points are referred to as *leaves*. An example schematic of a decision tree is shown in Figure 3.2.

The choice of observable and threshold applied are determined by minimising some *separation index*, such as GiniIndex G [75] or CrossEntropy E [76]. These are commonly defined in terms of the *purity* p, defined for binary classification as

$$p = \frac{s}{s+b}, \tag{3.12}$$

$$G = 2 \times p \times (1-p) = \frac{sb}{(s+b)^2}, \tag{3.13}$$

$$E = -p \times \ln(p) - (1-p) \times \ln(1-p), \tag{3.14}$$

where s and b are the number of signal and background events respectively. Consider a toy example in which there are six events, with the following observable values:

Classification	N_{jet}	N_{lep}
Signal	4	2
Signal	4	1
Signal	5	1
Background	4	1
Background	3	2
Background	3	2

The GiniIndex and CrossEntropy for the initial node are then

$$p = \frac{3}{3+3} = 0.5, \tag{3.15}$$
$$G = 2 \times 0.5 \times (1 - 0.5) = 0.5, \tag{3.16}$$
$$E = -0.5 \times \ln(0.5) - (1 - 0.5) \times \ln(1 - 0.5) = 0.693. \tag{3.17}$$

If a requirement is made that $N_{\text{lep}} < 2$, the GiniIndex and CrossEntropy, averaged over the resulting nodes, becomes

$$p_1 = \frac{1}{1+2} = 0.333 \quad , \quad p_2 = \frac{2}{2+1} = 0.666, \tag{3.18}$$
$$G = ((2 \times 0.333 \times (1 - 0.333)) + (2 \times 0.666 \times (1 - 0.666))/2$$
$$= 0.444, \tag{3.19}$$
$$E = (-0.333 \times \ln(0.333) - (1 - 0.333) \times \ln(1 - 0.333) +$$
$$-0.666 \times \ln(0.666) - (1 - 0.666) \times \ln(1 - 0.666))/2$$
$$= 0.637. \tag{3.20}$$

The separation index in both cases has thus reduced, and so this could be a candidate for the first splitting. However, if a requirement is instead applied at $N_{\text{jets}} \geq 4$

$$p_1 = \frac{3}{3+1} = 0.75 \quad , \quad p_2 = \frac{0}{0+2} = 0, \tag{3.21}$$
$$G = 0.09375, \tag{3.22}$$
$$E = 0.56233. \tag{3.23}$$

Now, both G and E reduce much more, and so this splitting is preferred to the first one tested. A decision tree will scan over all observables, find the optimal threshold for each, and choose the

one which results in the lowest value of the chosen separation index.

A single decision tree by itself is an example of a *weak* learner. If complexity is left unhindered, the tree will become overtrained. If complexity is limited, it it less susceptible to overtraining, though not immune, and performance will be limited: many leaves will not be completely pure in any particular class, and thus some events will be misclassified. *Boosting* in this context is an extension of the decision tree algorithm to improve the performance, by building an *ensemble* (or *forest*) of decision trees. In this algorithm, after the initial tree is built, those events which are misclassified are weighted up such that they carry more importance in a subsequent decision tree, changing the separation index score and thus building a different tree. The results of this tree are then similarly weighted, and this continues for some chosen number of iterations. The resulting output of the forest is the weighted performance of each (weak learning) tree, and so by combining many of these a BDT obtains better performance and is thus referred to as a *strong* learner. Examples of boosting algorithms include AdaBoost [69] and GradientBoost [77].

BDTs are very sensitive to overtraining, due to the precise manner in which the splittings are determined. There are two common strategies to minimise this overtraining: *bagging* and *pruning*. In a bagged BDT, only some (randomly chosen) subset of the training data is used per tree. By varying the exact events used per tree, the forest is less likely to be tuned precisely to the full training data. It also reduces the likelihood of a small number of outlier events being heavily boosted and becoming highly influential. *Pruning* is the process of removing the least important leaves after the full decision tree has been built. If removing a splitting node has less than some minimum effect on the output of the tree, it can be safely removed without a significant loss in performance, while making the tree more robust against overtraining.

A BDT has many tuneable hyperparameters. For example, one may choose to limit the maximum depth of the tree, the minimum size of a splittable node, or the number of trees in the forest. There is also the choice of the bagging fraction per tree, the pruning criterion, the choice of boosting method and associated coefficients, and the choice of separation index. There is no universally well performing set of hyperparameters, and these must be optimised on a case by case basis.

3.4.3 Deep Neural Networks

The term *neural network*, coined via loose analogy to the function of neurons in a brain, is an umbrella term for a family of MVA algorithms. In essence, a neural network approximates the true mapping $y = f^*(\boldsymbol{x})$ of a classification y given a set of features $\boldsymbol{x}$ with the function $f(\boldsymbol{x};\boldsymbol{\theta})$, by learning the parameters $\boldsymbol{\theta}$ that best describe the true mapping. This function f may be further broken down into *layers*, such that $y = f_1(f_2(f_3(\boldsymbol{x};\boldsymbol{\theta})))$, and neural networks which contain more than one such layer are known as *Deep Neural Networks*. Each layer consists of many *nodes*: the

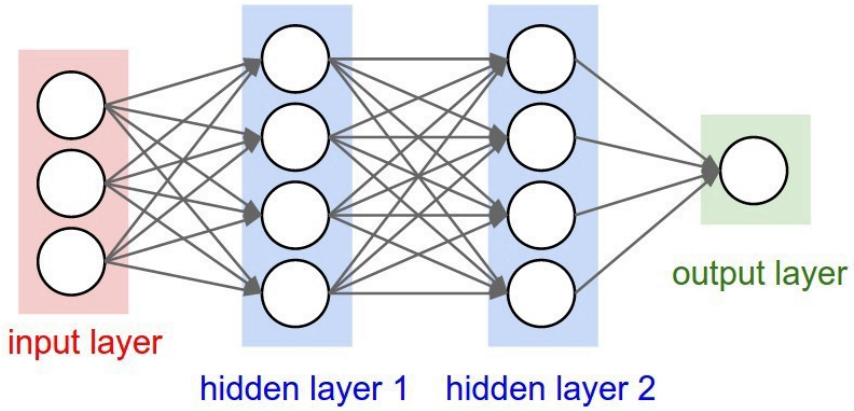


Figure 3.3: Model of a Deep Neural Network containing two hidden layers [78].

input layer has one node per input feature, while the output layer can have as many outputs as desired (just one for the case of binary classification, or one per class for multiclass output), and each *hidden* layer in between can have an arbitrary number. This structure is shown in Figure 3.3. The connections between each of these nodes are the associated parameters θ, called the *weights*, and if each node in a given layer is connected to each node in the subsequent layer, it is referred to as *fully connected*. No prior intuition of the problem is necessary up front when to initialise these weights: $\boldsymbol{\theta}$ can be initialised randomly, and learning happens via the updating of the weights after seeing some training data: a process known as *back propagation* [79].

The training of a DNN is a series of *forward passes* – calculating the network output for a given event – followed by the back propagation, where the network output is compared to the true value, and the weights are updated to better approximate the true value next time. The forward pass consists of simple linear algebra. First, the vector of inputs $\boldsymbol{x}$ is multiplied by the weights $\boldsymbol{\theta}$ connecting the first hidden layer, and all of the connections to each node in the hidden layer are summed. This sum S is then passed through the *activation function*, to calculate the node score. This node score is then analogous to an input feature from the input layer, and it is subsequently passed forward identically until reaching the output layer, where the node scores are finally the actual network output. The term hidden layer comes from the fact that the node scores for these layers are hidden from the user. An example of the forward pass phase of a simple network is shown in Figure 3.4.

The activation function is simply some mathematical function used to transform the inputs. Using non-linear functions, such as $f(x) = \tanh(x)$, allows the DNN to learn higher dimensional correlations between the inputs, and it is this introduction of non-linearity that provides much of the power of deep learning models. Other popular activation functions include the *sigmoid* $f(x) = \frac{1}{1+\exp^{-x}}$ and *rectified linear unit* (*ReLU*) $f(x) = \max(0, x)$, which are shown along with tanh in Figure 3.5.

Back propagation works backwards from the output layer, comparing the true label to the

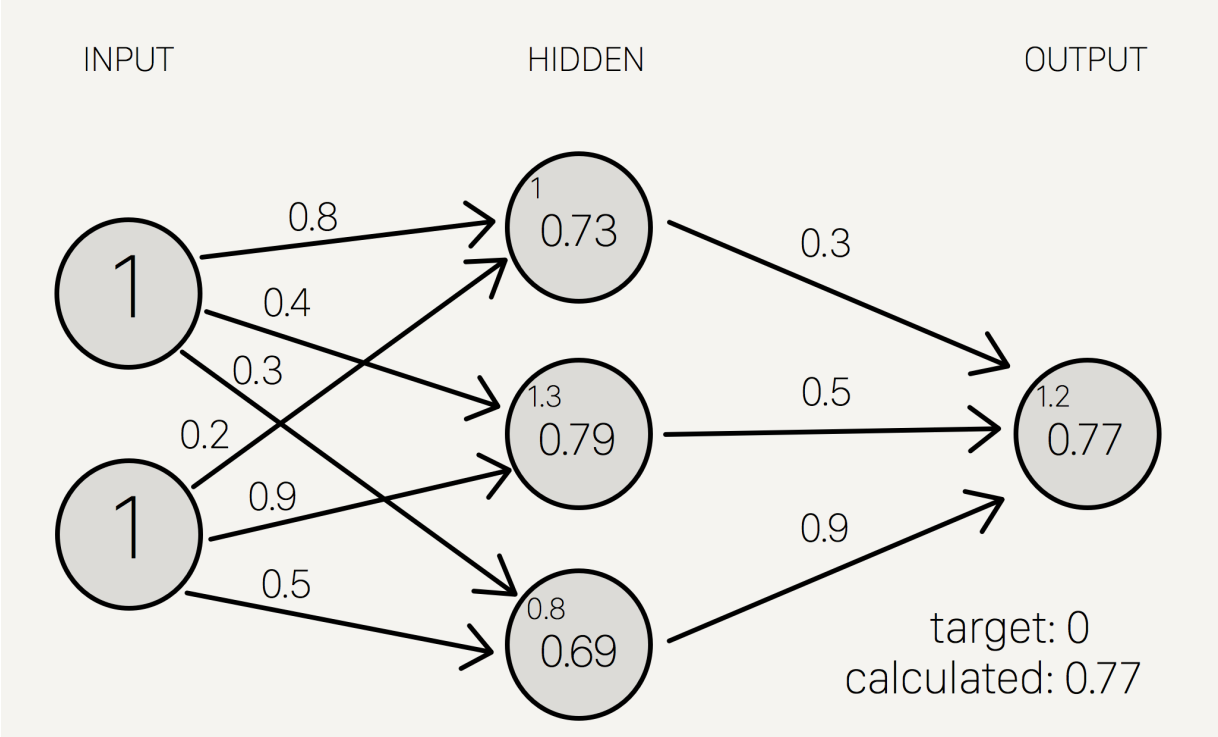


Figure 3.4: A toy example of a forward pass for a simple, fully connected network. The small number in each node of the hidden layer is the sum of the inputs × weights, and the larger number is the node score after passing through a sigmoid activation [80].

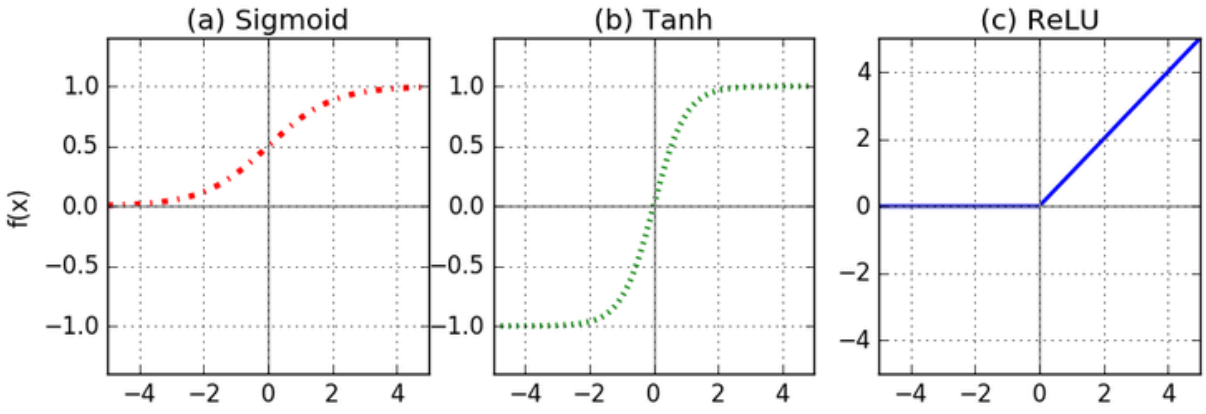


Figure 3.5: Examples of common non-linear activation functions [81].

prediction of the network output via the *loss function*. Common loss functions include the mean squared error E_{MSE} [82] and Cross-Entropy E_{CE} [76], given by

$$E_{\text{MSE}} = \frac{1}{2}(y_{\text{true}} - y_{\text{out}})^2, \tag{3.24}$$

$$E_{\text{CE}} = -(y_{\text{true}} \log(y_{\text{out}}) + (1 - y_{\text{true}}) \log(1 - y_{\text{out}}), \tag{3.25}$$

with y_{true} the true classification and y_{out} the network output. The network objective – to act as the best approximation to the true mapping between the inputs and true label – can be easily re-expressed as being to minimise the loss function for all output nodes. Thus, it is natural to think in terms of gradients, and calculate derivatives of the error E of the output o with respect to the weight θ used

$$\frac{\partial E}{\partial \theta} = \frac{\partial E}{\partial o} \times \frac{\partial o}{\partial S} \times \frac{\partial S}{\partial \theta}, \tag{3.26}$$

where we have taken advantage of the chain rule, with S the sum of the inputs prior to activation. If the loss function chosen is differentiable, $\frac{\partial E}{\partial o}$ is easily calculable. Similarly, if the activation function is differentiable, then $\frac{\partial o}{\partial S}$ can be calculated. The final piece, $\frac{\partial S}{\partial \theta}$, is the simple derivative of the vector multiplication and summing. To make this clear, consider the concrete example of mean squared error and sigmoid activation with a single output node. Then

$$E = \frac{1}{2}(\text{true} - o)^2 \tag{3.27}$$

$$\implies \frac{\partial E}{\partial o} = (\text{true} - o), \tag{3.28}$$

$$o = \frac{1}{1 + \exp^{-S}} = \frac{e^S}{1 + e^S} \tag{3.29}$$

$$\implies \frac{\partial o}{\partial S} = \frac{e^S}{(1 + e^S)^2} = o(1 - o), \tag{3.30}$$

$$S = O \cdot \theta \tag{3.31}$$

$$\implies \frac{\partial S}{\partial \theta} = O. \tag{3.32}$$

Here, O is the node output from the previous layer. Thus, by continuing this use of the chain rule all the way back to the original inputs, the dependence of the output error on each weight can be calculated. To decrease the error, the weights are updated by subtracting this gradient, thus moving towards the minimum of the loss function. This process is known as *gradient descent*. In practice, the value subtracted is multiplied by an additional factor called the *learning rate*, which is usually $<< 1$. In this way, the weights change by a smaller amount at each step, which may prevent them from oscillating around a minimum. Of course, too small a learning rate may get stuck in a local minima, and/or drastically slow the convergence – and so this is a hyperparameter

of the network that must be chosen. Since calculating these gradients takes time, it is common to forward propagate a number of events together, calculate the error of these, and then update the weights using this error. The hyperparameter for the number of events per update is the *batch size*. When all of the data has been passed through the network, this is an *epoch*, and the number of these is another hyperparameter. Similar to a BDT, a DNN can be sensitive to overtraining, and good choices of these and other hyperparameters is important. Other strategies such as *dropout*, where connections between nodes are randomly dropped for each batch, can also help to prevent overfitting. The architecture of the network can itself be considered as a set of hyperparameters, with the number of layers, the number of nodes in each of those layers, the amount of dropout, and the choice of activation/loss functions all important choices with no a-priori favoured values.

A subtle consideration when using neural networks is data preprocessing. Unlike a BDT, which performs linear selections on each input, the absolute scale of the input variables is important, as made clear by substituting Equation 3.32 into Equation 3.26. Thus, inputting two variables of widely varied scales – for example, the p_T of a jet in MeV might be $O(10^6)$, while the η of a jet is $O(1)$ – would give vastly different relative importance to the larger variable. To get around this, the data should be normalised such that all inputs are similar in magnitude and variance, ensuring robust behaviour in the gradient descent. For example, a simple method to do this would be to rescale all of the inputs to be centred at 0, and to normalise by the mean of the distribution to have unit variance.

In addition, there are many other deep learning architectures useful in particle physics beyond fully connected layers, including (but not limited to) convolutional nets [83], LSTMs [84], or generative adversarial networks [85], which have not been used in the analyses presented and so are not discussed here. An example application of a fully connected DNN, used in the boosted $t\bar{t}H, H \to b\bar{b}$ analysis for jet tagging, is presented in Section 9.3.

Chapter 4

ATLAS and the LHC

We are stuck with technology when what we really want is just stuff that works.

– Douglas Adams

Lying at an average depth of 100 m below the surface of the French and Swiss countryside, the Large Hadron Collider (LHC) [86] is the worlds largest piece of scientific apparatus built to date. It is designed to collide protons (and heavy ions) at energies of up to 14 TeV, to study the particles and forces of nature in conditions equivalent to those less than 1 billionth of a second after the Big Bang. Superconducting magnets operating at 7.7 T bend bunches of protons around the 27 km ring, with RF cavities accelerating them to 99.999999% of the speed of light. The two beams are brought together once every 25 ns at one of 4 locations around the ring, where the ATLAS [87], CMS [88], LHCb [89], and ALICE [90] experiments are each housed. LHCb and ALICE are designed specifically to study heavy flavor physics in pp collisions and quark-gluon plasma physics in heavy ion collisions respectively, while ATLAS and CMS are general purpose particle detectors built to study the physics of the Standard Model as well as search for new physics beyond the Standard Model. The studies presented in this document have all been performed using data collected with the ATLAS detector, which is described in detail in Section 4.2.

4.1 The Large Hadron Collider

The LHC is housed in the same tunnel that was previously home to the Large Electron-Positron Collider (LEP) [91]. The tunnel was originally dug between 1983 and 1988, with the two ends eventually meeting with an error of just 1 cm. At its deepest point, the tunnel is 175 m below the surface, owing to its location at the foot of the Jura mountains, and is built at a gradient of 1.4% in order to minimise excavation costs as well as line up with the other tunnels in the CERN accelerator complex. Protons are fed to the LHC via this network, depicted in Figure 4.1, from a humble beginning in a small bottle of hydrogen gas. The hydrogen is subsequently ionised

and accelerated to 50 MeV using a linear accelerator known as LINAC2[1], which has been in operation since 1978. The protons subsequently enter the BOOSTER, in use since 1972, where they are accelerated up to 1.4 GeV before being passed to the Proton Synchotron (PS), which first accelerated particles in 1959. When they leave the PS, the protons have reached 25 GeV, and are then fed to the Super Proton Synchotron (SPS). The SPS, itself 7 km in circumference, dates back to 1976 and was the location for the discovery of the $W^{\pm}$ [92, 93] and Z [94, 95] bosons, which earned Carlo Rubbia and Simon van der Meer the 1984 Nobel Prize for Physics. When the protons leave the SPS and enter the LHC, they have reached an energy of 450 GeV and traversed more than half a century of history and technology. While the rest of the accelerator complex uses traditional magnet technology, in order to reach energies in the TeV range the LHC relies on superconducting magnets. Each of the 1232 15 m long dipole magnets, weighing 35 tonnes each, must be cooled to just 1.9K to achieve resistance free electrical conduction. Quadropole magnets are used to squeeze the beams at each interaction point, narrowing the effective size of the beams to maximise number of colliding particles.

4.1.1 Operation

The first operational run of the LHC (*Run 1*) took place between 2009 and 2013, with proton collisions at centre of mass energies lower than the design energy of $\sqrt{s}$ = 14 TeV, at $\sqrt{s}$ = 900 GeV, $\sqrt{s}$ = 7 TeV, and $\sqrt{s}$ = 8 TeV. After a two year shutdown for maintenance and upgrade work, the second operational run (*Run 2*), began in 2015 and continued until the end of 2018. This time, collisions took place at $\sqrt{s}$ = 13 TeV, chosen as a compromise between energy reach, accelerator operational stability, and the time required for magnet preparation. Protons are circulated around the machine in *bunches*, with a fixed number of protons per bunch and a set spacing between them. The *instantaneous luminosity* $\mathcal{L}$ of the machine is a measure of the number of interactions taking place at any given moment in the machine, and is dependent only on beam parameters such as these, as well as the transverse size of the beam. Assuming two identical beams, the equation for luminosity is

$$\mathcal{L} = \frac{kN^2}{4\pi s a_h a_v}, \tag{4.1}$$

where s is the bunch spacing in units of time, k is the number of bunches per beam, N is the number of protons per bunch in the two beams, and a_h and a_v are the effective beam sizes in the horizontal and vertical directions [97]. The total dataset size is then the integrated luminosity over the data taking period, $\mathcal{L}_{\text{int}} = \int \mathcal{L} \mathrm{d}t$, often measured in units of inverse femtobarns (fb), where 1 fb = 10^{-39}cm^2.

[1]LINAC2 will be decommissioned during the second long shut down of the LHC, with the new LINAC4 accelerator, designed to reach higher luminosity, replacing it as the first step in the proton accelerator chain for Run 3.

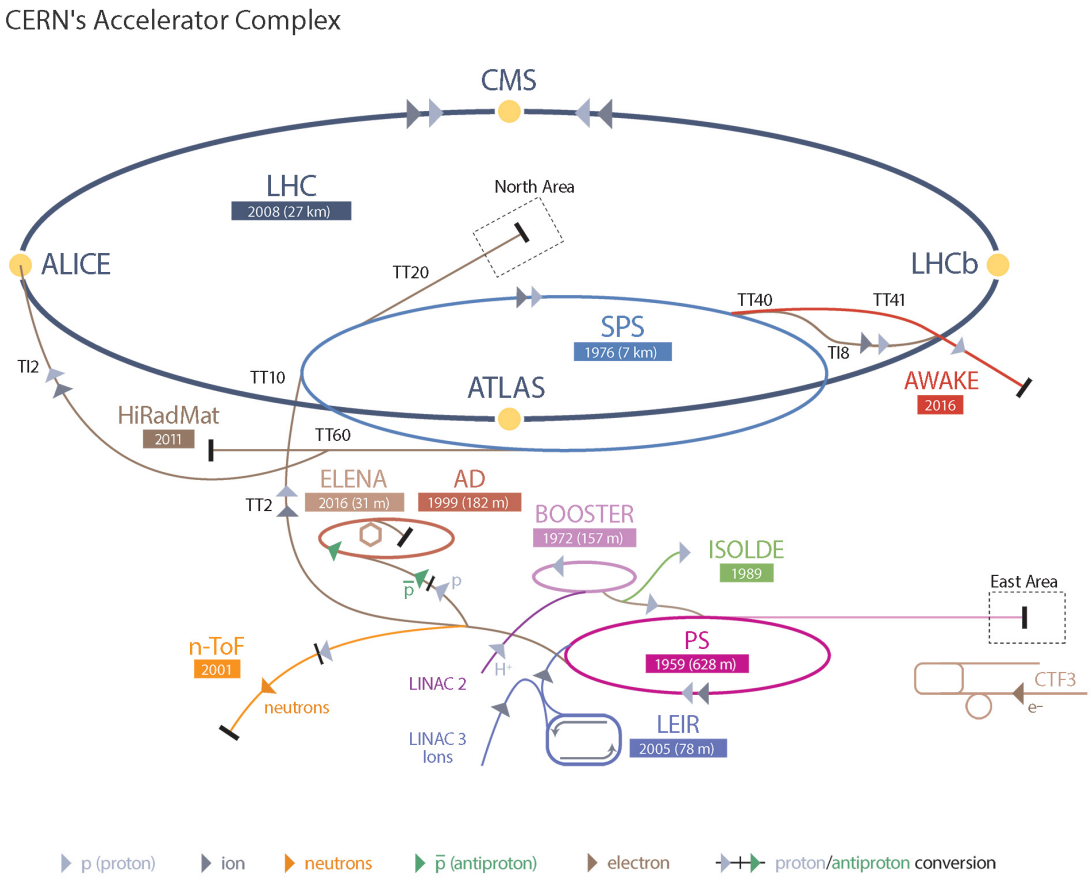


Figure 4.1: Diagram showing the CERN accelerator complex [96]. The LHC itself is shown in dark blue.

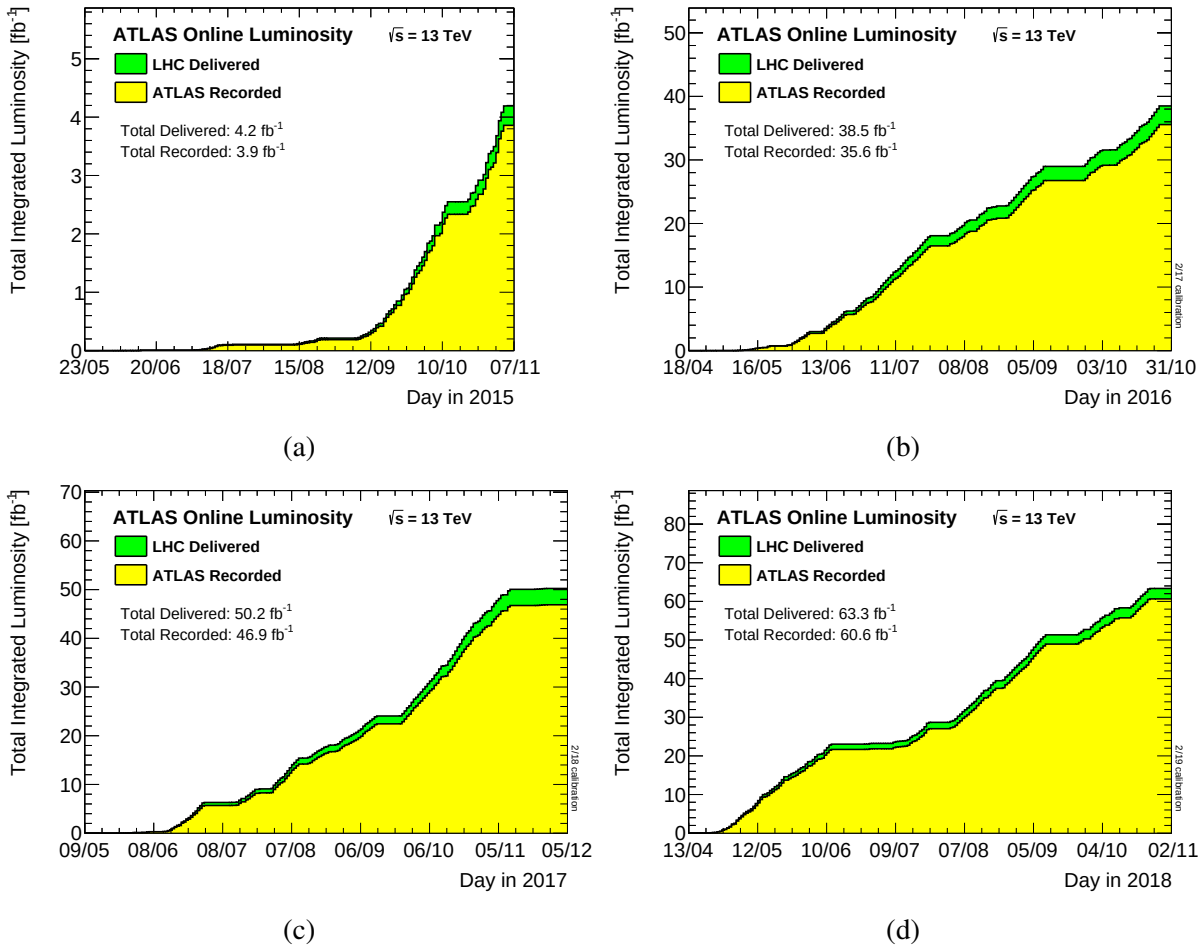


Figure 4.2: The luminosity delivered by the LHC and recorded by ATLAS through each year of Run 2 data taking, for (a) 2015 (b) 2016 (c) 2017 (d) 2018 [99]. The green area is the total integrated luminosity delivered by the LHC, beginning from the moment stable beams are declared until they are lost. The yellow area represents the data recorded by ATLAS, with the inefficiency relative to the green primarily due to data acquisition inefficiency and the time taken for ATLAS to ramp up after stable beams are declared.

The analyses performed here utilise data collected by the ATLAS detector during Run 2 of the LHC, at a centre of mass energy of $\sqrt{s} = 13$ TeV. The amount of data collected each year is shown in Figure 4.2. In total, 156 fb^{-1}was delivered by the LHC, of which 147fb^{-1}was recorded by ATLAS. Of this, 139 fb^{-1} was collected with all subsystems of the ATLAS detector operational. This information is summarised in Figure 4.3. Full details of the LHC performance during Run 2 is summarised in [98].

When bunches cross in the centre of ATLAS, there is almost always more than one visible pp collision per crossing[2]. This is known as *pileup* (μ). With higher instantaneous luminosity comes larger μ, making it more difficult to reconstruct the hard interactions of interest in the midst of many other concurrent pp collisions. Originally the LHC was designed to reach a peak

[2]With the exception of some special, low-μ runs performed in 2017 and 2018, demonstrated by the small peak at $\mu = 1$ in Figure 4.4.

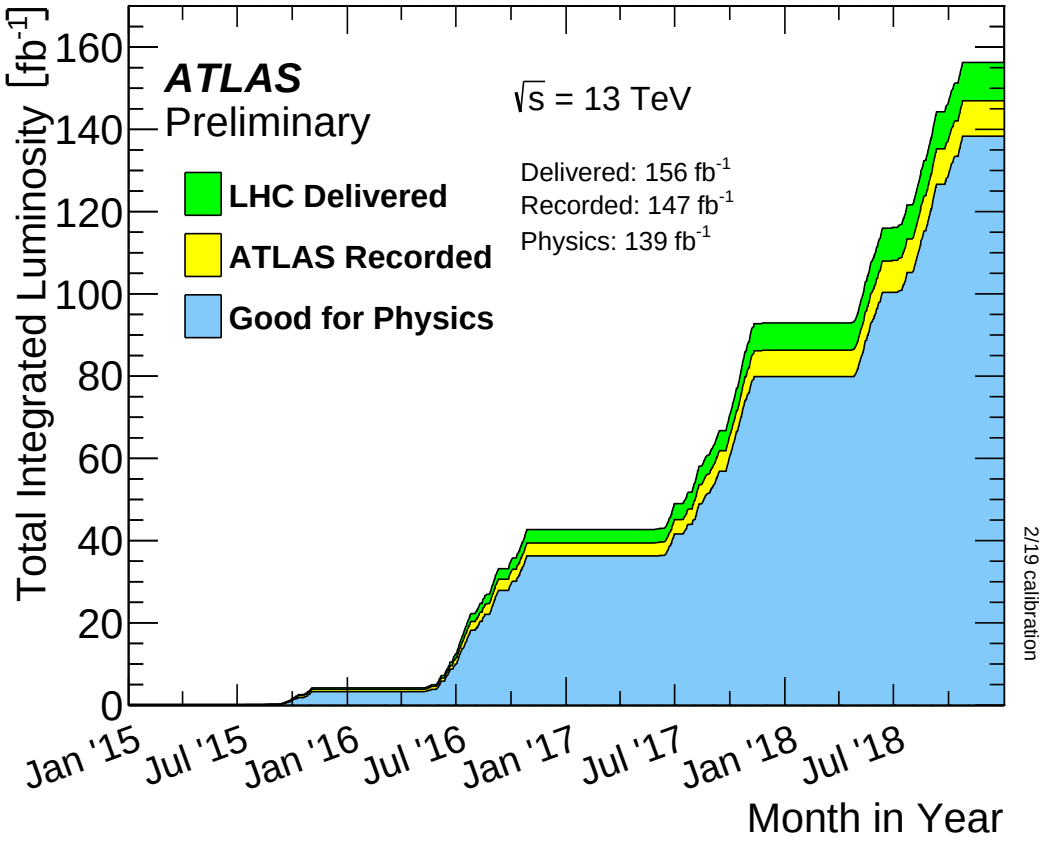


Figure 4.3: The total luminosity delivered and recorded over the whole of Run 2 by the ATLAS detector [99]. The yellow and green areas are the same as those in Figure 4.2. The blue area, labelled "Good for Physics", represents the data collected with all ATLAS subsystems fully operational.

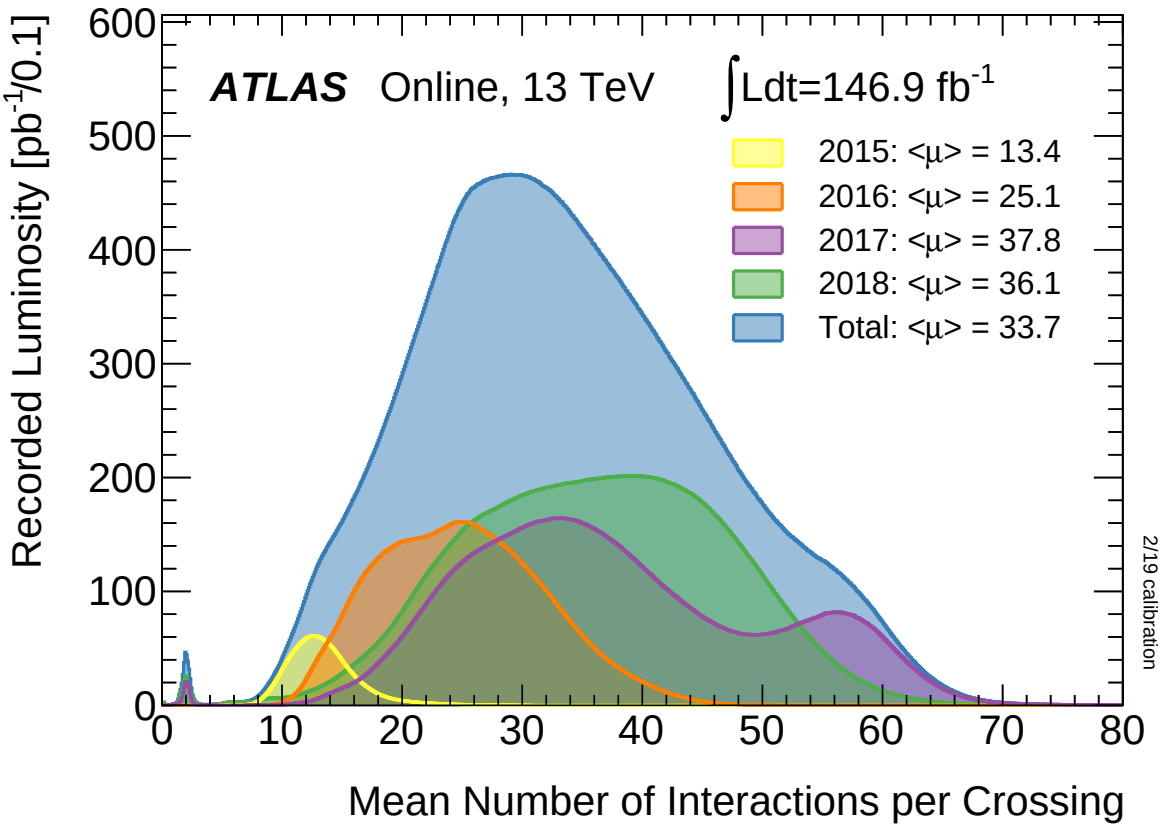


Figure 4.4: The average number of visible *pp* interactions per bunch crossing in ATLAS, per year and total [99].

luminosity of 10^{34} cm^{-2}s^{-1}, though this was exceeded already by 2016, with a record high of 2.1×10^{34} cm^{-2}s^{-1} achieved in 2018. This has meant an average pile up $\langle\mu\rangle = 33.7$ over the whole of Run 2, as shown in Figure 4.4. In the latter half of 2016, a pile up of $\langle\mu\rangle = 60$ was common, after the LHC fill scheme had to be changed to account for *"the gruffalo"* – persistent problems in a particular sector of the accelerator that caused regular beam dumps. The original design value was $\langle\mu\rangle = 19$. Higher pileup means more activity in the detector, making reconstruction of physics objects trickier, and identification of which objects originate from the hard process more difficult. Efficiently handling high pileup is increasingly important, in particular for Run 4 and beyond, when the LHC enters the *High Luminosity LHC* (HL-LHC) phase [100].

4.2 ATLAS

A Toiroidal **L**HC **A**pparatu**S** may be a tenuous acronym at best, but the ATLAS detector [87] is also the largest general purpose particle detector ever built, at 25 m in diameter and 44 m in length, as shown in Figure 4.5. Built in the shape of a barrel around the beam pipe at Point 1 of the LHC tunnel, ATLAS combines a multitude of detector technologies to effectively identify

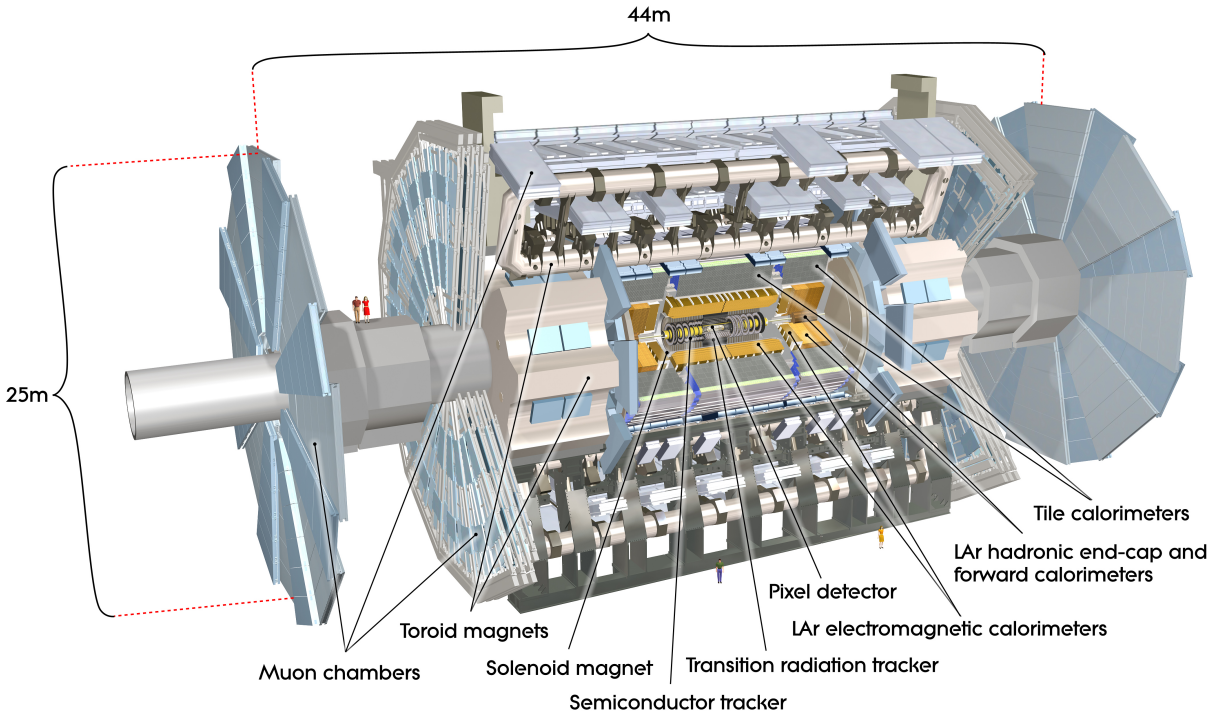


Figure 4.5: The ATLAS Detector. The scale of the machine is demonstrated by the two small figures standing between the muon wheels.

almost all of the particles in the Standard Model, with the exception of the neutrinos, which escape without interaction. The barrel structure allows for these technologies to be combined in a layered manner, with each stage being useful for the detection of different types of particles, as shown in Figure 4.6. Much like the particles themselves, the following sections will travel through each layer in turn, beginning at the centre of the detector and moving steadily outwards.

This thesis will use a right-handed coordinate system, with z the beam direction, ϕ the azimuthal angle and θ the polar angle. Generally θ is replaced in analyses in favour of the quantity *rapidity*, $y = 0.5 \ln\left(\frac{E+p_z}{E-p_z}\right)$, or *pseudorapidity* $\eta = -\ln\tan(\theta/2)$, where E is the particle energy and p_z is the momentum in the beam direction. The quantities y and η are often used interchangeably.

4.2.1 Inner Detector

The detector sub-system closest to the beam pipe is the *Inner Detector* (ID) [102]. The ID is used to construct high resolution tracks of charged particles, which can be used to measure particle momentum (in combination with the magnet system) as well as primary and secondary vertices associated with particle decays. To do this, three distinct technologies are used – the Pixel detector, the Semiconductor Tracker, and the Transition Radiation Tracker – in a layered structure, depicted in Figure 4.7.

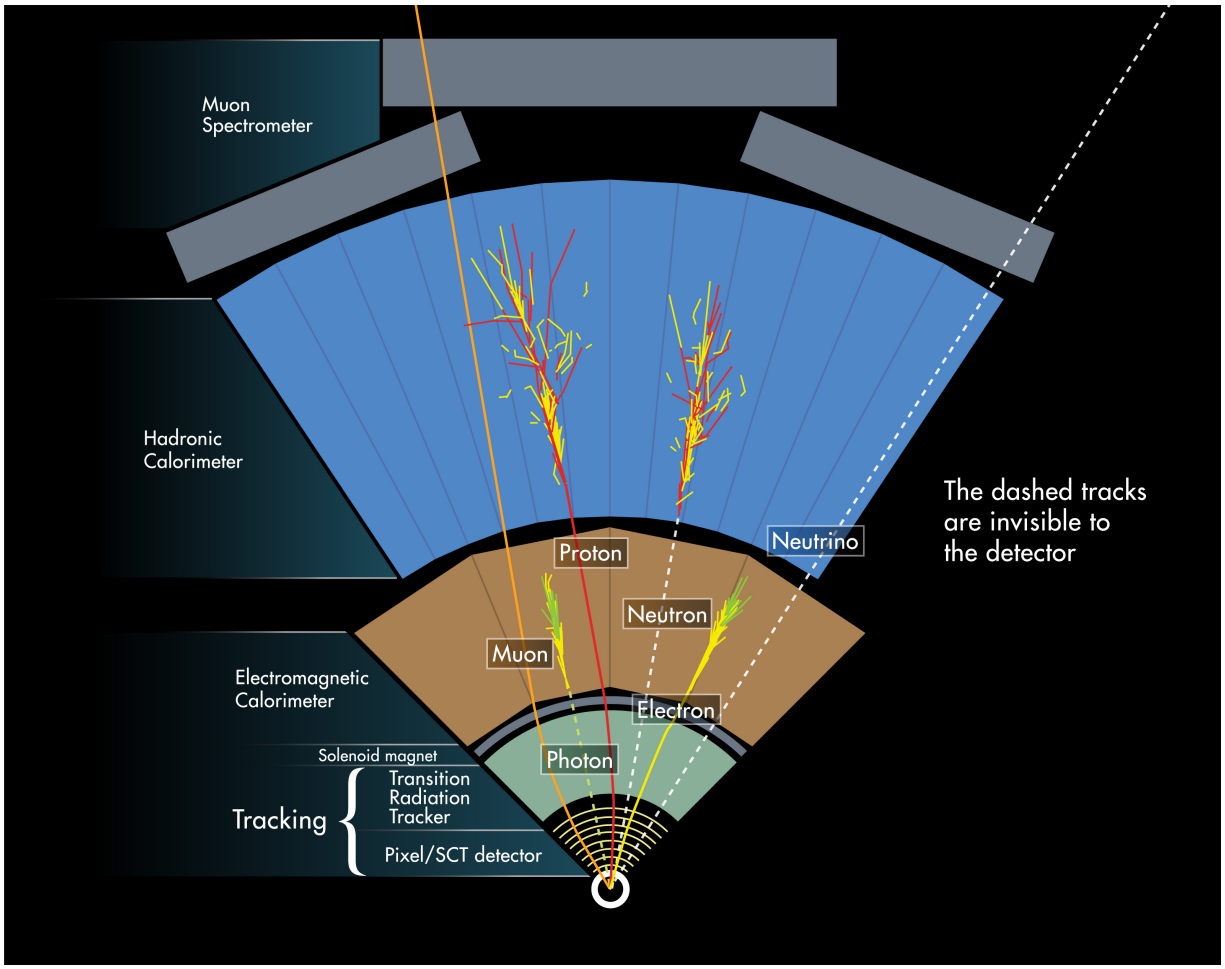


Figure 4.6: Cartoon demonstrating how different types of particle interact with each layer of the ATLAS detector [101]. Each type of particle has a different combination of signatures in the subsystems.

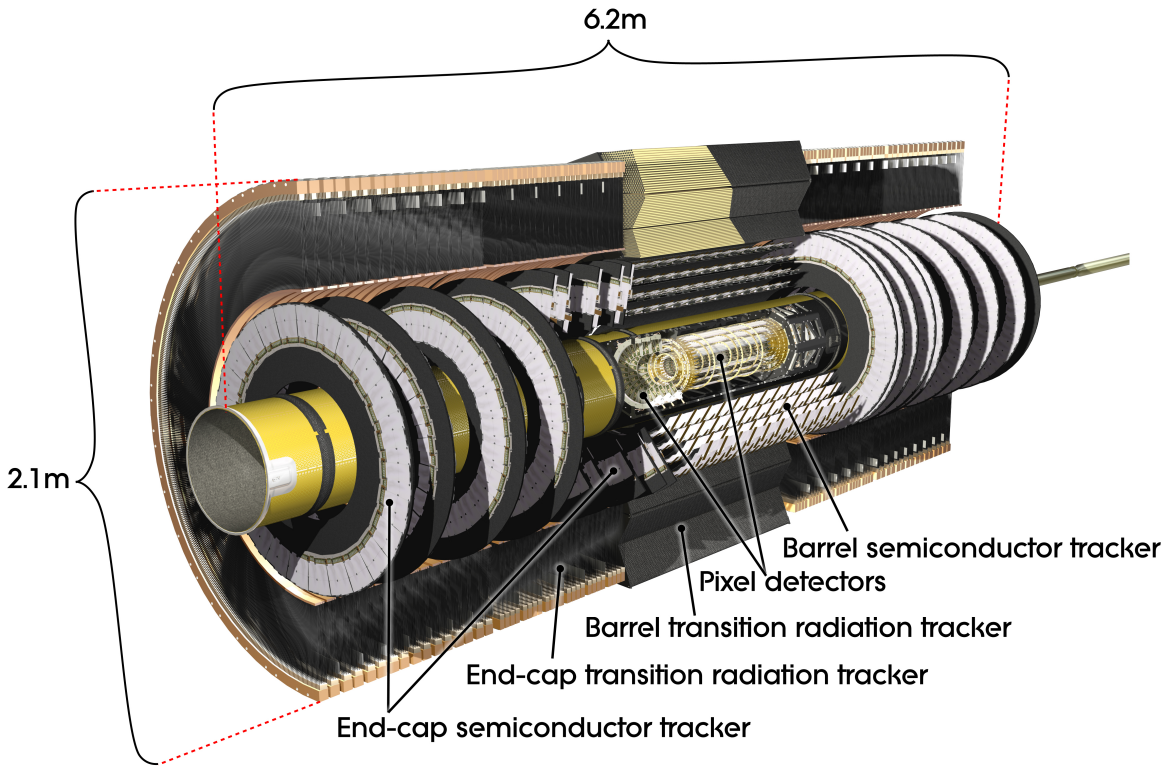


Figure 4.7: The layout of the ATLAS inner detector [103].

Pixel Detector

Closest to the beam pipe is the Pixel detector [104], itself composed of four layers in the barrel and three layers in the endcap[3], including the Inner B-Layer (IBL) [106]. The IBL was installed during the shutdown between Run 1 and Run 2, and is the closest piece of ATLAS hardware to the LHC beam. The Pixel detector is a system of 1968 silicon semiconductor sensors providing coverage in the region $|\eta| < 2.5$. In total, there are 80.4 million pixels, of size 50×250 μm^2 in the IBL and 50×400 μm^2 in the other layers, providing spatial resolution of 10 μm (8 μm) in the plane perpendicular to the beam direction and 115 μm (75 μm) in the beam direction itself in the 3 primary layers (IBL).

Semiconductor Tracker

The Semiconductor Tracker (SCT) [107], next in line after the Pixel detector, uses similar silicon sensors this time arranged in strips. There are a total of 15912 sensors in the SCT covering the region $|\eta| < 2.5$, with 4 barrel layers and 9 disks in each endcap region. The 6 cm long strips are arranged at a stereo angle of 40 mrad between neighbouring sensors, providing resolution of 17 μm in the transverse and 580 μm in the longitudinal directions.

[3]It is turtles all the way down [105].

Transition Radiation Tracker

Unlike the other ID layers, the Transition Radiation Tracker (TRT) [108] uses straw tube sensors 4 mm in diameter, providing coverage up to $|\eta| < 2.0$. The TRT has a resolution of 130 μm in the transverse plane, and no longitudinal measurement. The TRT does however provide around 36 hits per track, having 73 straw planes parallel to the beampipe in the barrel and 160 radial straw planes in the endcap. The gas mixture used is 70% Xenon, 27% CO_2 and 3% Oxygen, and by using two different thresholds, hits from transition radiation can be be discriminated from tracking hits to enhance electron identification.

4.2.2 Calorimeters

Outside the tracking detectors are the calorimeters [109], which measure the energies of charged and neutral particles. Metal plates known as *absorbers* are used to halt the incident particles, causing a shower of energy that is then measured with *sensing elements*. There are a total of four calorimeter systems in ATLAS. The LAr Electromagnetic Calorimeter is designed to detect particles which primarily interact electromagnetically, such as electrons and photons, with a fine granularity. The remaining systems are used to measure the hadronic particles that are not fully absorbed by the first layer. The full system is designed to be thick enough to prevent almost all punch-through, with the exception of the minimum ionising muons, which are subsequently detected in the dedicated muon systems. The layout of the calorimeter system is shown in Figure 4.8.

LAr Electromagnetic Calorimeter

The liquid argon electromagnetic (EM) calorimeter [111] is the first detector layer after the ID, and uses lead plate absorbers and liquid argon sensing material. Since the central solenoid magnet (described in Section 4.2.4) is placed between the ID and the calorimeters, the EM calorimeter shares a vacuum vessel with the solenoid to minimise the number of vacuum walls. The EM calorimeter is composed of a three layer barrel ($|\eta| < 1.475$) and two-stage endcap (three layers for $1.375 < |\eta| < 2.5$ and two layers for $2.5 < |\eta| < 3.2$), providing the highest precision in the $|\eta| < 2.5$ region that is also covered by the silicon tracking detectors. There is also a presampler layer for $|\eta| < 1.8$ which measures the energy lost by electrons and photons before entering the calorimeter. The exact granularity in $\Delta\eta \times \Delta\phi$ varies with $|\eta|$, but is generally around 0.025×0.025 in the region of the tracking acceptance and 0.1×0.1 in the more forward region.

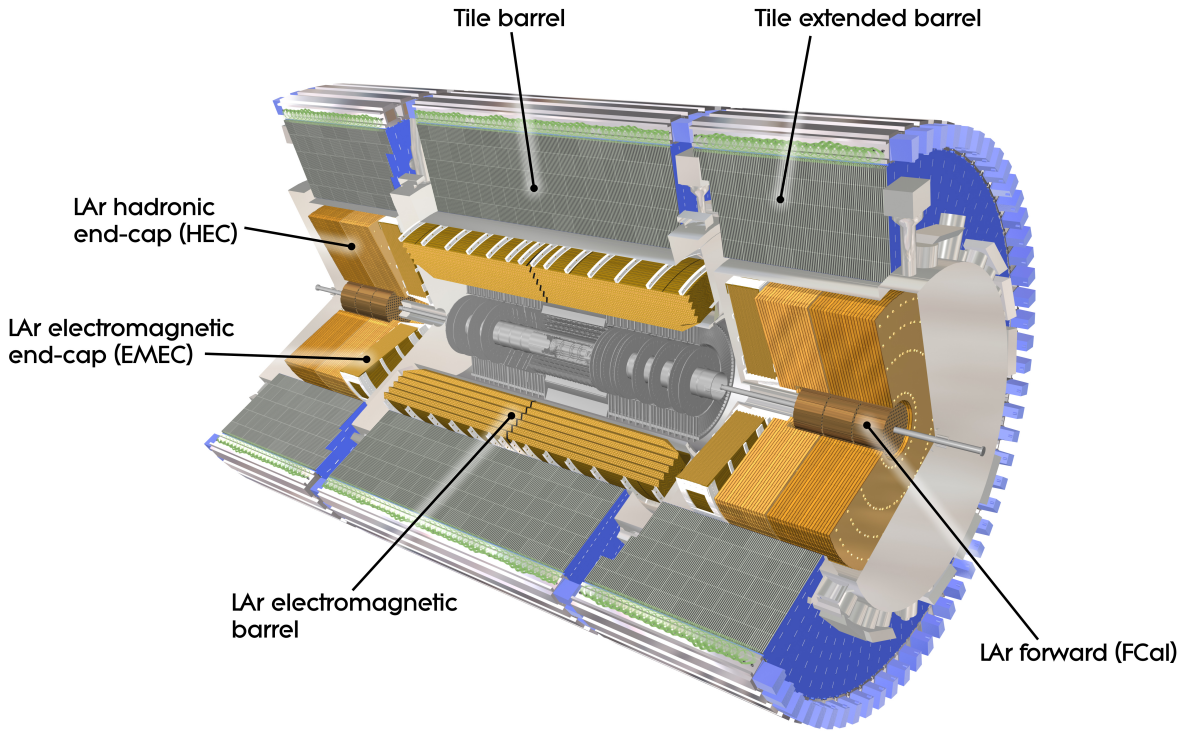


Figure 4.8: The ATLAS calorimeter system [110].

Tile Hadronic Calorimeter

Beyond the EM calorimeter is the Tile calorimeter [112], consisting of the barrel ($|\eta| < 1.0$) and extended barrel ($0.8 < |\eta| < 1.7$). Both of these regions use steel absorbers and scintillating tile sensing elements, with three segmented layers, providing a detector thickness of 7.4 interaction lengths to prevent punch-through. The Tile has a granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$ in the first two layers and 0.2×0.1 in the final layer.

LAr Hadronic Endcap Calorimeter

The Tile does not extend to the endcap region, with instead the Hadronic Endcap Calorimeter (HEC) [111] a further lead-LAr calorimeter sharing the endcap cryostat of the EM calorimeter. This covers the region $1.5 < |\eta| < 3.2$, to slightly overlap with the the tile at low η and the forward calorimeter at high η. The HEC is made up of two, 32 segment wheels per endcap, which are further divided into two layers for a total of four. For $1.5 < |\eta| < 2.5$ the granularity is 0.1×0.1, compared to 0.2×0.2 in $2.5 < |\eta| < 3.2$.

LAr Forward Calorimeter

In order to extend the calorimeter coverage as far as $|\eta| < 4.9$, there is also a LAr forward calorimeter (FCal) [111] in the endcap cryostats. Due to the higher radiation levels in the forward region, the FCal must be separated from the EM calorimeter by around 1.2 m in the z direction, and thus the punch-through prevention requirement of at least 10 interaction lengths necessitate a high-density design. The three module system uses copper absorbers in the first layer for electromagnetic sensitivity and tungsten in the subsequent layers for hadronic activity, with thin layers of liquid argon sensing material. The granularity is approximately 0.15×0.15 at $|\eta| = 3.5$ and 0.3×0.3 at $|\eta| = 4.5$.

4.2.3 Muon Spectrometer

The final layer of the ATLAS onion is the muon spectrometer [113], which is designed to measure the momentum of charged particles which exit from the calorimeter systems. Primarily, such particles will be muons, which do not deposit much of their energy in the calorimeters. There are four different technologies used for this purpose: Monitored Drift Tubes (*MDT*), Cathode Strip Chambers (*CSC*), Resistive Plate Chambers (*RPC*), and Thin Gap Chambers (*TGC*). Together these provide the necessary capabilities of both precision tracking (primarily MDT and CSC) and fast tracking for triggering (RPC and TGC). A cut away view of the muon systems is shown in Figure 4.9, with a schematic view also shown in Figure 4.10.

Monitored Drift Tubes

Monitored Drift Tube chambers [116], made up of 3-8 layers of 30 mm tubes, provide most of the precision tracking coverage. MDT chambers cover the area up to $|\eta| < 2.7$, with the exception of the innermost layer of the endcap, which covers only the region $|\eta| < 2.0$. There are three layers of MDT chambers in the barrel, located on and inbetween the toroid magnet coils. In the endcap are wheel shaped chambers with 16 overlapping large and small segments to provide full coverage. They are located both in front, above, and behind the endcap toroid, as well as a further outer wheel approximately 7.5 m further out, as shown in Figure 4.10. The endcap layer above the toroid was extended between Run 1 and Run 2 in order to improve resolution in this region. The layout is designed such that a straight track should traverse at least three different chambers, with small exceptions due to support structures and access points, with each chamber measuring to a precision of approximately 35 μm.

Cathode Strip Chambers

The innermost region of the first endcap chamber is composed of Cathode Strip Chambers [117] instead of MDTs, due to the need for a higher rate and increased time resolution in the forward

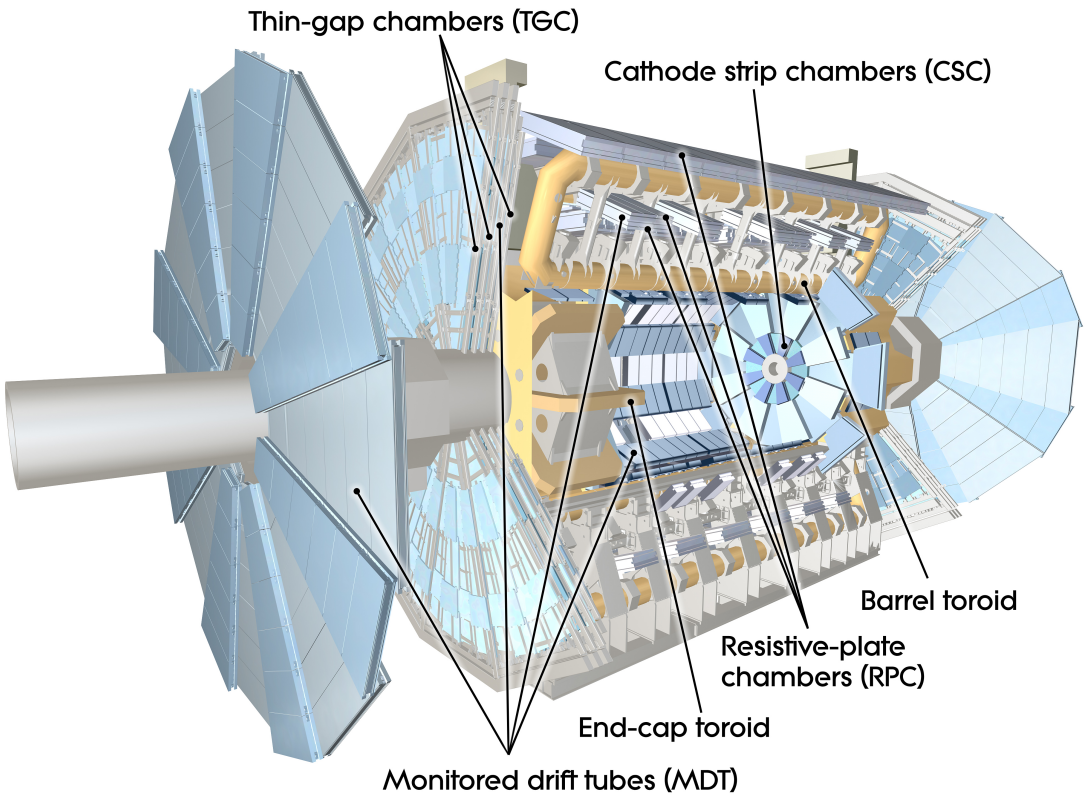


Figure 4.9: The ATLAS muon and magnet systems [114].

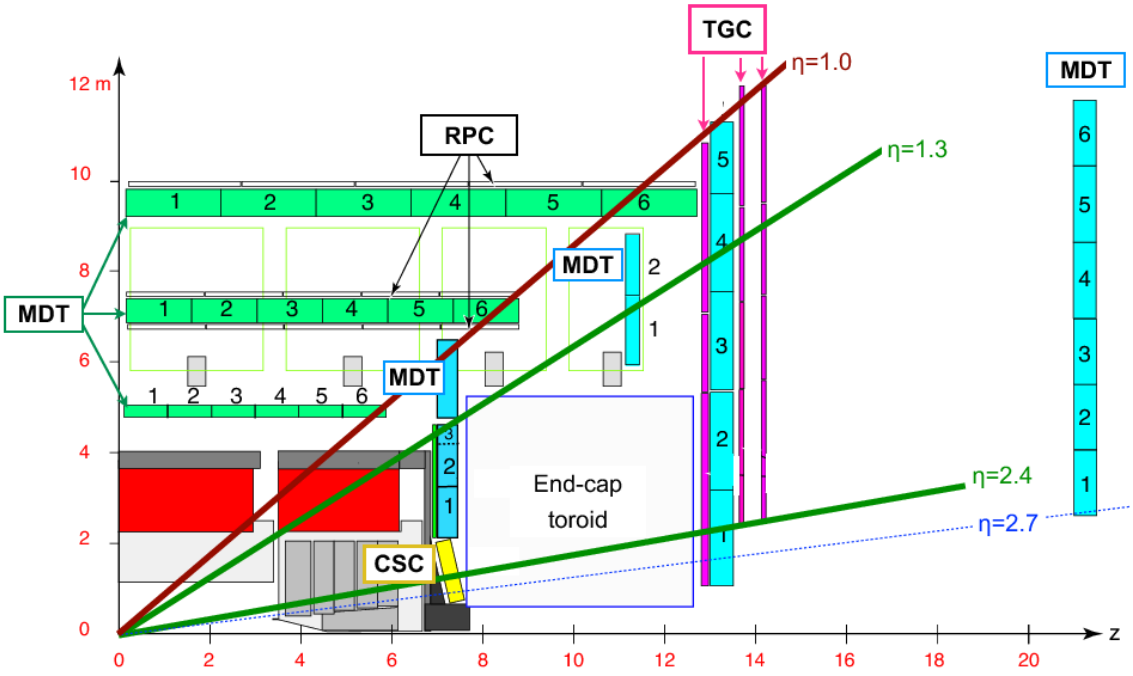


Figure 4.10: Schematic view of the muon spectrometer [115].

region. A similar overlapping large and small segment layout, as used in the MDTs, ensures full coverage and a resolution of 40 μm in the bending plane.

Resistive Plate Chambers

In order to efficiently trigger on muons, there must also be a fast tracking system in the muon spectrometer. To do this, Resistive Plate Chambers [118] are used in the barrel region, $|\eta| < 1.05$. These have a response time of 15-25 ns and time resolution of 1.5 ns. These are located on the outside of the MDTs, on both sides of the middle layer and outer side of the outer layer. Each of these have two independent layers providing both η and ϕ information, giving up to six measurements in both coordinates. This measurement in the non-bending plane is also used in tracking to complement the MDTs, which do not measure in this orthogonal plane.

Thin Gap Chambers

For the endcap region, $1.05 < |\eta| < 2.4$, the trigger and non-bending plane responsibilities are covered by Thin Gap Chambers. These have a similar response time as the RPCs, and resolution of 4 ns. There are 7 layers of TGCs surrounding the middle endcap MDT chamber, and a further two around the inner layer. As well as for triggering, the TGCs are used to extrapolate the azimuthal coordinate in the outer MDT chamber, which is made possible by the lack of magnetic field between these two chambers.

4.2.4 Magnets

The ATLAS detector incorporates four large superconducting magnets [119] – one solenoid magnet surrounding the inner tracking detector components, and three toroid magnets (one barrel toroid plus one in each endcap) which bend charged particles in the muon spectrometer. The magnet geometry of the full detector is shown in Figure 4.11.

Solenoid

The 2.5 m diameter Solenoid provides a 2 T magnetic field to the ID, bending charged particles that pass through the tracking detectors. It is cooled to a temperature of approximately 4.5 K to achieve superconduction using liquid helium. The magnet is 10 cm thick and 5.8 m long, with a total weight of 5.7 tonnes.

Barrel and Endcap Toroids

The Barrel Toroid magnet consists of eight coils, each contained in their own vacuum vessel and located outside the calorimeter systems. They are similarly cooled to 4.5 K with liquid helium,

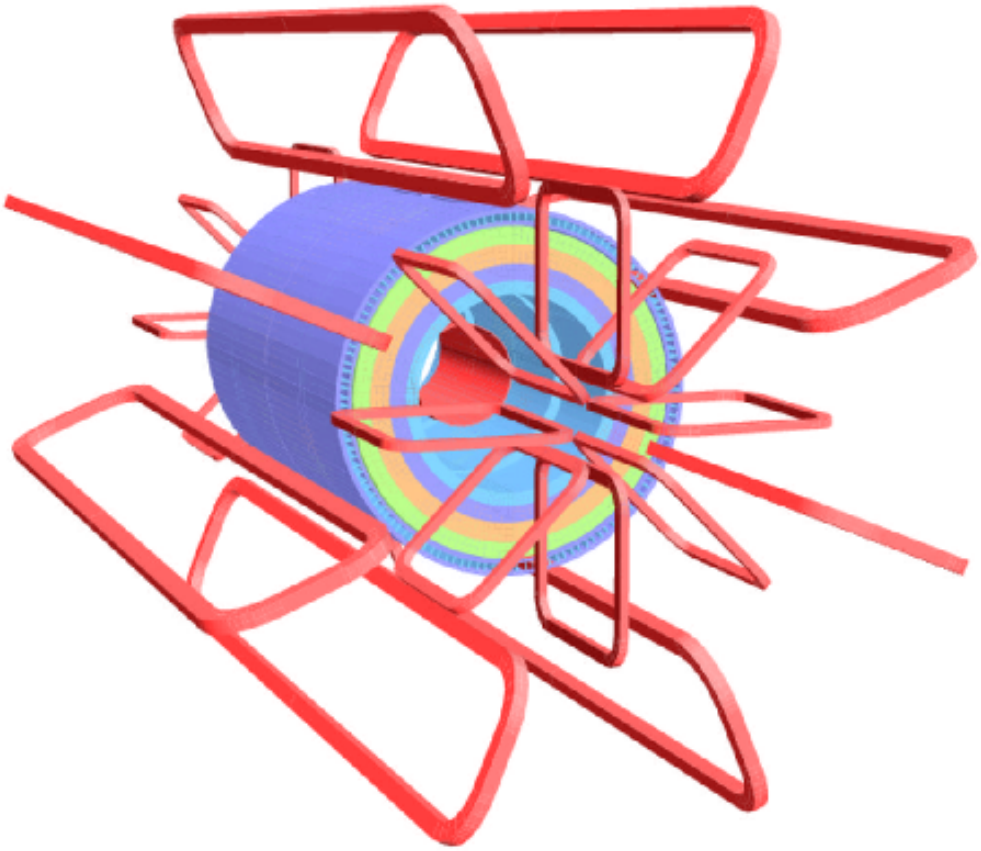


Figure 4.11: The geometry of the ATLAS magnet system [87], shown in red. The central solenoid provides a 2 T magnetic field to the ID, with the four calorimeter layers outside this, shown in purple, orange, and green. There are eight coils of the toroid magnet system in the barrel and and eight in each end-cap, which bend charged particles escaping the calorimeters and entering the muon system.

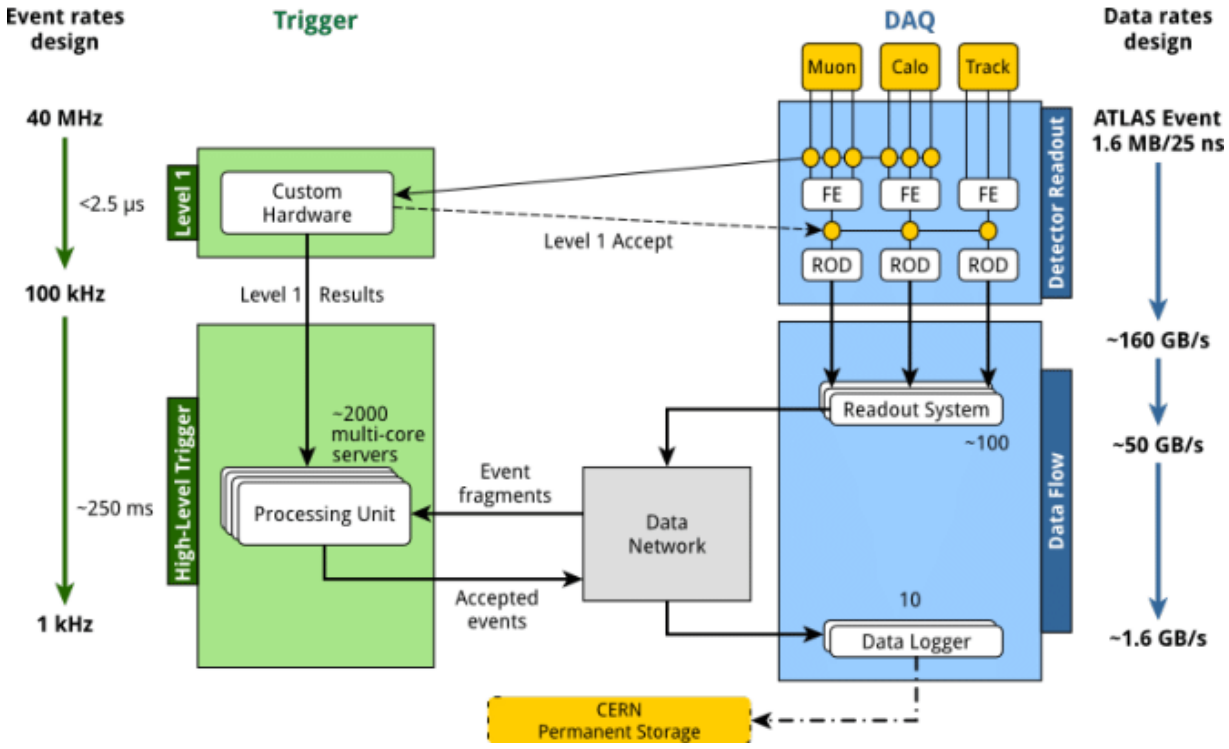


Figure 4.12: Schematic of the ATLAS TDAQ system, as used in Run 2 [120].

and have an inner and outer diameter of 9.4 and 20.1 m respectively, at a length of 25.3 m. The magnetic field varies from 0.15 – 2.5 T, averaging around 0.5 T, in the barrel. In each of the endcaps is a 10.7×5.0 m toroid, also consisting of eight coils and operating at 4.5 K. The magnetic field varies from 0.2 to 3.5 T in the endcap region.

4.2.5 Trigger

With bunches crossing at the heart of the ATLAS detector every 25 ns, the potential amount of data to record is overwhelming. Only a subset of this 40 MHz rate can thus be stored, requiring the use of a *trigger* which selects the subset of events to store based on some predefined criteria. The ATLAS trigger and data acquisition uses a two stage trigger to reduce the rate. These are referred to as the Level 1 (*L1*) trigger and the High Level Trigger (*HLT*), and are shown schematically in Figure 4.12.

Level 1 Trigger

Reducing the rate from the initial 40 MHz down to around 100 kHz, the L1 trigger [121] uses custom hardware to achieve a decision time of less than 2.5 µs. It utilises information from the calorimeter system and the RPC and TGC components of the muon spectrometer in order to identify high p_T objects such as muons, electrons/photons, jets, and hadronically decaying τ-leptons, as well as for events with large missing transverse energy or total transverse energy.

Events containing such objects are then passed to the HLT should they meet the L1 criteria. Since these objects are reconstructed quickly, the requirements at L1 are relatively loose, with a typical single electron or muon requirement at L1 corresponding to $p_T > 20$ GeV.

High Level Trigger

After passing at L1, the HLT trigger [122] uses a large farm of CPU servers to reconstruct the event with a greater precision than the L1. Information from the full detector is available in the HLT, though some information – such as ID tracking – is slower to process than others, and so the reconstruction is performed in *chains* of functions which utilise increasingly complex algorithms. At each stage an event may be rejected if it does not pass a set criteria, avoiding the need to run the later, more CPU intensive algorithms unless they are necessary. For example, a typical single muon chain will rebuild the candidate muon by matching the spectrometer information to crude ID tracking, subsequently running more precise tracking should the candidate be well reconstructed and rejecting it if not. A typical single muon chain will have a final $p_T > 26$ GeV in Run 2, though this was lower ($p_T > 24$ GeV) in the lower luminosity data taking in 2015.

In all of the analyses presented, single muon and electron triggers are used. The trigger p_T requirement subsequently impacts the event selection at analysis level, which is typically ~ 1 GeV higher than the trigger requirement to ensure full efficiency. The analysis in Chapter 7 uses only 2015 data, and thus the lepton p_T requirement is set to $p_T > 25$ GeV. The analyses in Chapters 8 and 9 utilise data from subsequent years in which the threshold was increased to 26 GeV, and thus the lepton p_T requirement is set to $p_T > 27$ GeV in those. The performance of the triggers is evaluated in both data and simulation, and any difference in response is applied to the simulated events as an event weight in order to correct it to match the data.

Fast Tracker

The Fast TracKer (FTK) [123] is a phase-1 upgrade to the ATLAS trigger system, presently being installed in preparation for Run 3 of the LHC, that will provide tracking information at the L1 output rate. Since tracking is a computationally expensive algorithm, information from the tracking detectors is not currently available at L1. With the high luminosity conditions that are planned for Run 3, this information can be useful to reduce trigger rates and keep thresholds low, and additional offers new triggering opportunities on signatures such as displaced tracks that can be useful in BSM searches. This is achieved with new FPGA hardware utilising massively parallel processing, utilising custom built pattern banks to fit tracks with $p_T > 1$ GeV. First, 8 of the 12 layers of the silicon tracking detectors are used, and a linear fit is performed to these hits using pre-calculated constants based on the detector geometry. This fit is subsequently extrapolated to the final 4 layers, and tracks surviving a quality threshold can then be passed to the HLT. This linear fitting procedure provides a helix parameter that is almost as good as

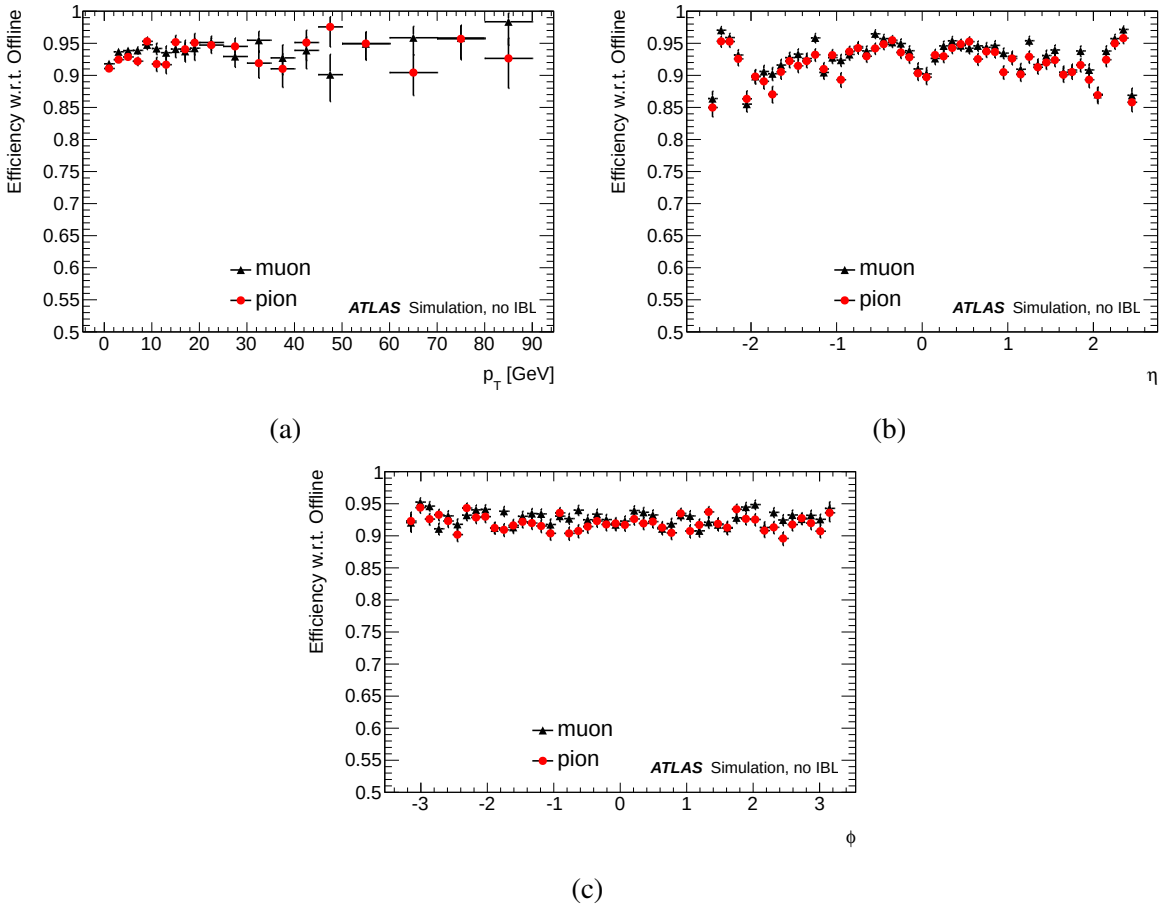


Figure 4.13: FTK tracking efficiency relative to standard offline tracking in ATLAS, for both muons (black triangles) and pions (red circles) [123].

a helical fit, while being far less CPU intensive. Parallelisation allows the huge number of possible track paths to be processed simultaneously, and the combination of these two allows a tracking efficiency of approximately 90% for both muons and charged pions with respect to offline tracking, as shown in Figure 4.13. This efficiency is too low to replace the standard ID tracking completely, but still allows for many new applications of tracking information in the trigger algorithms. The full system is designed to provide full detector tracking information at 100 kHz, matching the L1 output rate.

4.2.6 FTK Integration in Muon Trigger Chain

Keeping the single lepton trigger thresholds as low as possible is crucially important to the physics programme of ATLAS, particularly as these triggers are the most commonly used in the collaboration. To do this, the CPU overhead of the single lepton chains must be kept as small as possible. The FTK upgrade provides additional information in these chains that can be utilised for this purpose. One such example is the single muon trigger chain depicted in Figure 4.14.

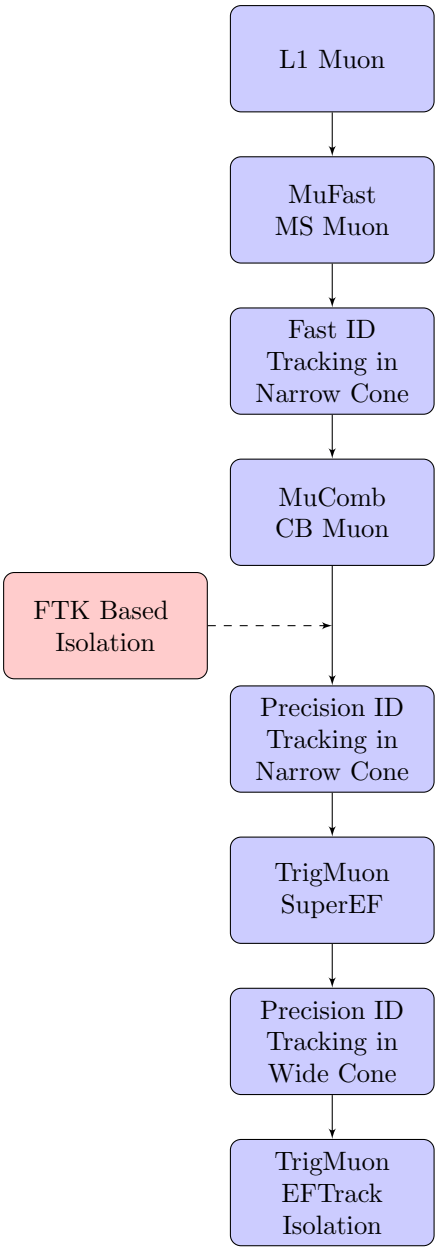


Figure 4.14: Flow chart depicting the single muon trigger chain used in ATLAS. The existing algorithm is shown in blue, with the red box marked "FTK Based Isolation" indicating the new algorithm that can be added using FTK information.

In this chain, tracking information is used in two ways: to build the muon (both using fast ID tracking and precision ID tracking), and to check the isolation of the muon (using precision ID tracking only). This isolation is important to reject muon candidates which result from heavy hadron decays, rather than those from $W^{\pm}$ and Z bosons, which are of primary interest. The FTK efficiency of $\sim 90\%$ is insufficient to replace any of these steps completely, but FTK tracks can be used to perform an earlier estimate of isolation, thereby not requiring the precision ID tracking in events in which the muon is obviously insufficiently isolated. This should be optimised to save as much CPU as possible, without rejecting any events that would have been saved by the existing chain.

The existing isolation algorithm performs tracking in a cone of $R = 0.2$ around the muon candidate, summing the p_T of the tracks in this region (subtracting the p_T of the muon itself, in order to remove the track used in the muon building), and rejecting events based on the ratio $p_T^{\text{cone}}/p_T^{\mu}$. In Run 2, the isolation criterion using ID tracks was $p_T^{\text{cone}}/p_T^{\mu} < 0.12$. The new algorithm based on FTK tracks uses a similar strategy, with the only exception being the subtraction of the muon candidate p_T. Since the muon is built using ID tracks, the subtraction is now ambiguous – sometimes the FTK will not find the track that was used to build the muon candidate, and even when it is, the p_T measurement is often very different. This has the potential for rejecting an event that would have been saved by the precision isolation, if the FTK track had significantly larger p_T than the corresponding ID track. Two alternative strategies were tested: subtracting the track closest to the muon candidate in ΔR, and subtracting the leading track p_T. The two demonstrated similar performance, and so the conservative choice of subtracting the leading track p_T was chosen. The effect of this is demonstrated in Figure 4.15. With the strategy fixed, the optimal working point was found using the equations

$$\epsilon_1 = \frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{FF}}}, \tag{4.2}$$

$$\epsilon_2 = \frac{N_{\text{PP}}}{N_{\text{FP}} + N_{\text{FF}}}. \tag{4.3}$$

Here, the first and second subscripts represent the algorithm using ID tracks and FTK tracks respectively, with P indicating that the muon passed and F indicating that the muon failed the isolation criteria. Thus, N_{PP} represents the number of events in which the muon passed both isolation criteria, N_{FF} represents the number of events which in which the muon failed both isolation criteria, and N_{FP} represents the number of events in which the muon passed the ID track isolation but failed the FTK track isolation. The optimal working point for the cone isolation is thus where $\epsilon_2 \simeq 1$, and ϵ_1 is minimised under this requirement. The two efficiencies as a function of isolation variable $p_T^{\text{cone}}/p_T^{\mu}$ are shown in Figure 4.16 for events from a $t\bar{t}$ sample containing both prompt and non-prompt muons from the W and B-Hadron decays respectively. It is observed that the same value of 0.12 is optimal also for the FTK based isolation. This choice of working point

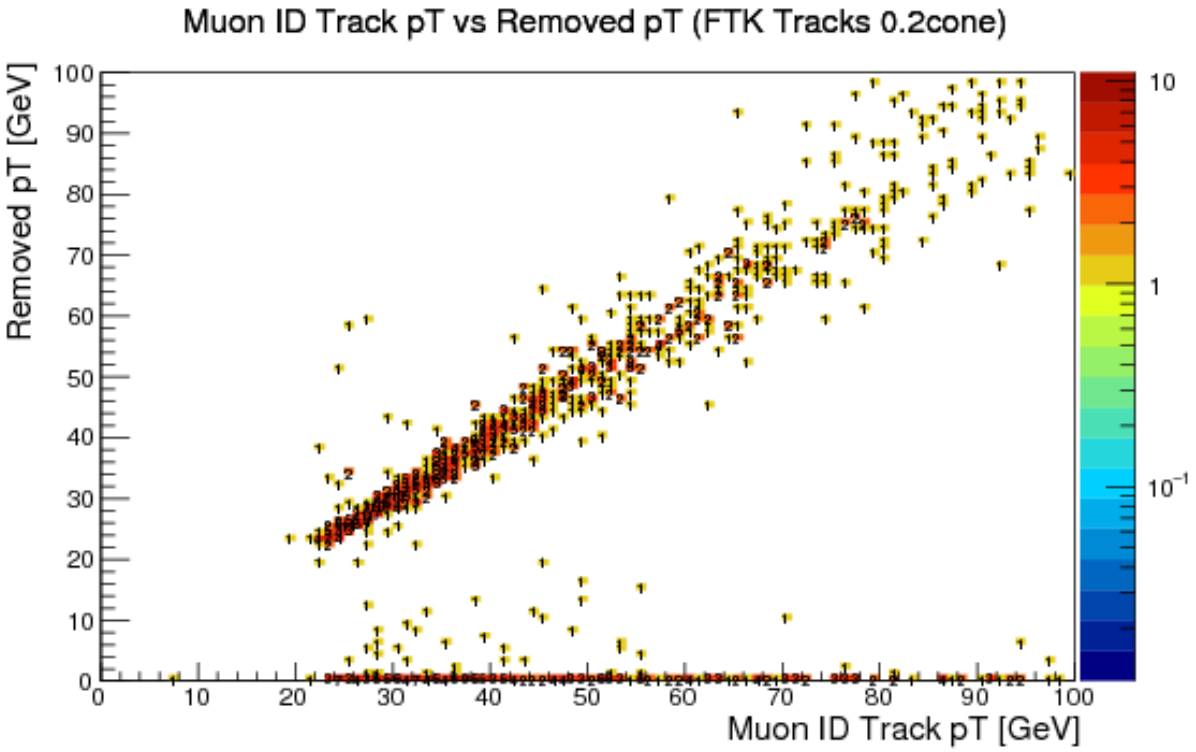
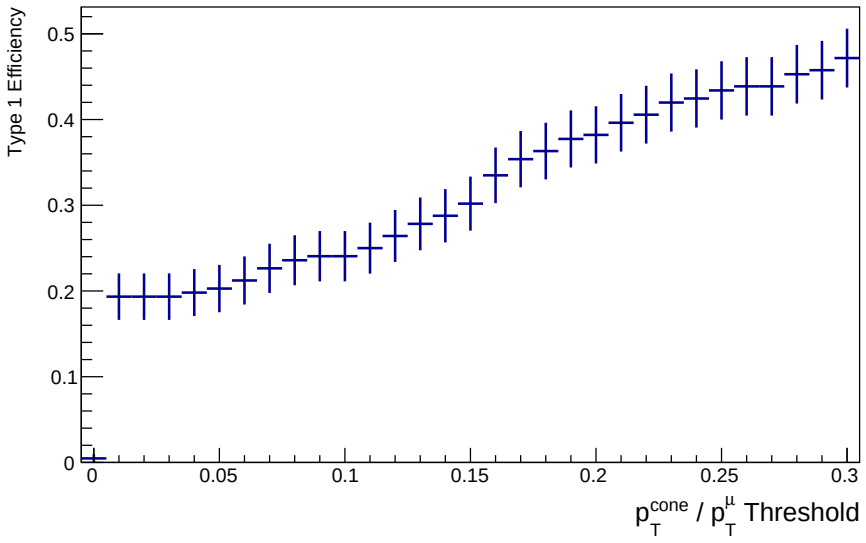
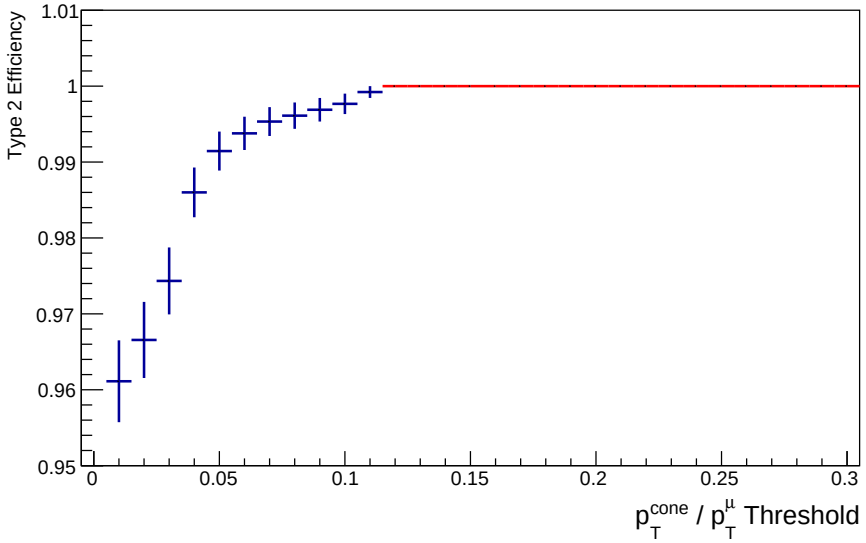


Figure 4.15: 2D plot of the Muon candidate ID track p_T vs the removed track p_T in the FTK isolation algorithm, in a sample of simulated $t\bar{t}$ events. Those events which lie at 0 on the y-axis correspond to events in which no tracks, including the muon track itself, were reconstructed by the FTK. Since this means that the event will not be rejected by the algorithm, this does not present a problem. Good diagonality is otherwise observed.

was validated using a tag and probe approach [124] with simulated $Z \rightarrow \mu^+\mu^-$ events, shown in Figure 4.17. The trigger efficiency including the FTK based isolation is approximately 100% with respect to the nominal chain. Further cross checks were performed in real data collected using enhanced bias triggers, which select events containing high p_T L1 seeds at random, with simulated FTK tracks. These checks are shown in Figure 4.18, demonstrating the type of events that are being removed by the FTK isolation by showing the event properties at the precision ID isolation stage. As expected, the rejected events are busy, non-isolated events with high track multiplicity and poorly reconstructed muons, with no bias in the muon candidate kinematics. The CPU time taken to process these events is shown in Figure 4.19, demonstrating an 8% reduction in the average CPU time when including the FTK isolation in the trigger chain. This reduction will be important in the next run of the LHC should the FTK be successfully installed, keeping the thresholds of the single muon triggers – on which the analyses presented all rely – as low as possible.



(a)



(b)

Figure 4.16: The (a) ϵ_1 (type 1) and (b) ϵ_2 (type 2) efficiencies for the isolation algorithms, as defined in Equations 4.2 and 4.3.

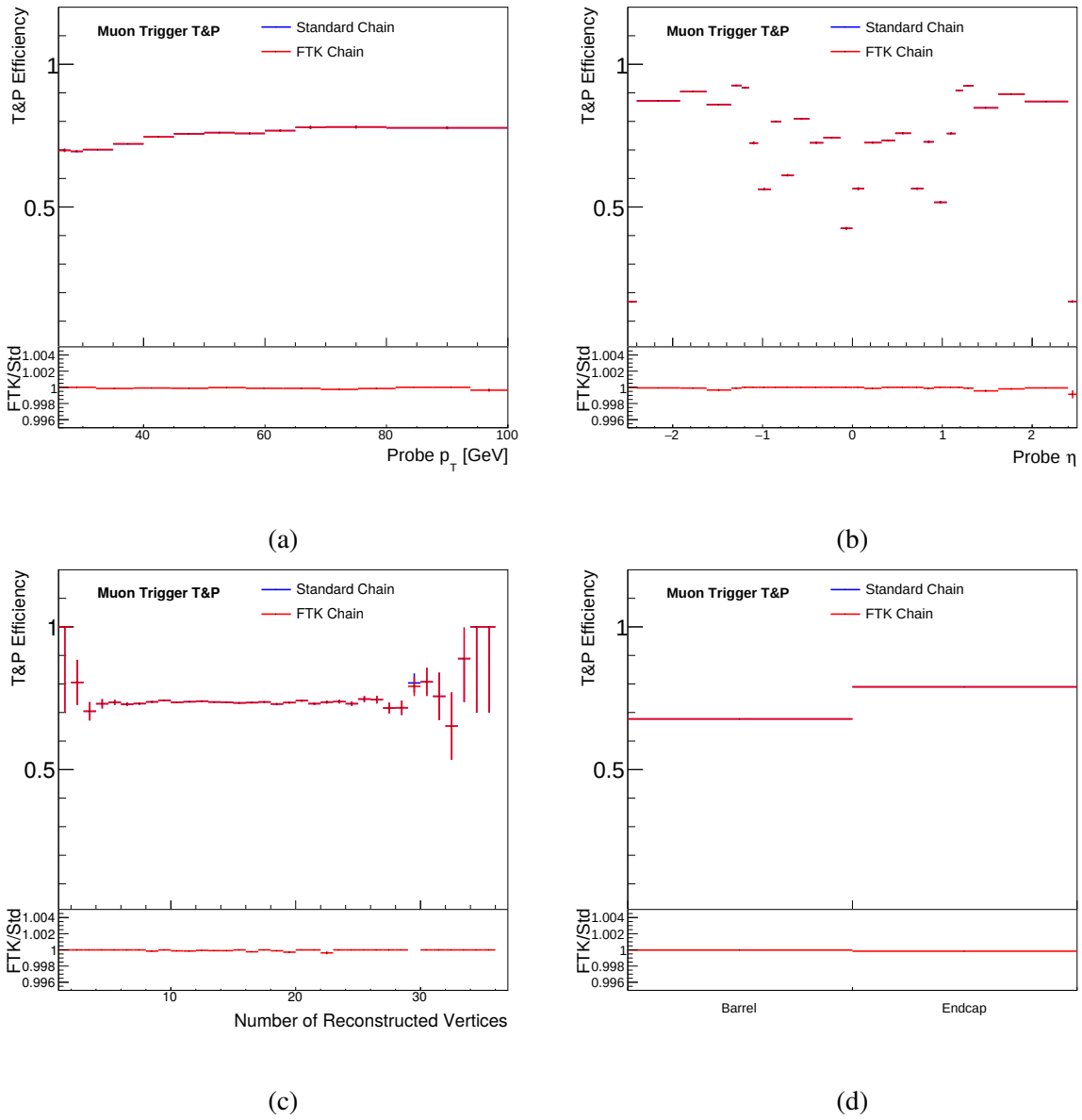


Figure 4.17: The tag and probe efficiencies for the standard single muon chain and the FTK isolation chain. The efficiency loss with the FTK is chain is at the sub per-mille level.

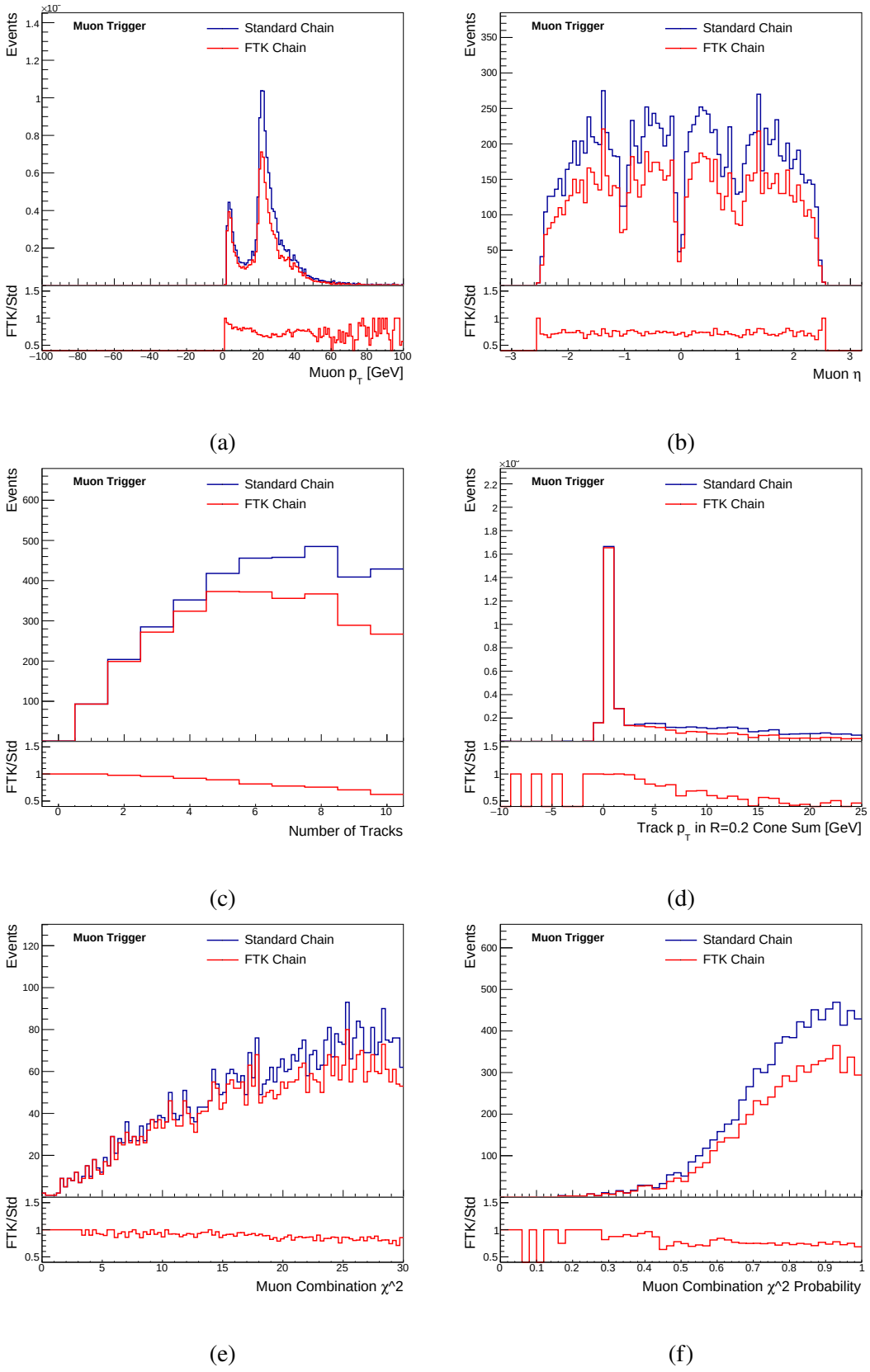


Figure 4.18: The control distributions for the precision ID tracking algorithm, both with (blue) and without (red) the FTK based isolation in the chain. As expected, it is events which high track activity and poor muon reconstruction which are removed, with no clear bias in muon kinematics.

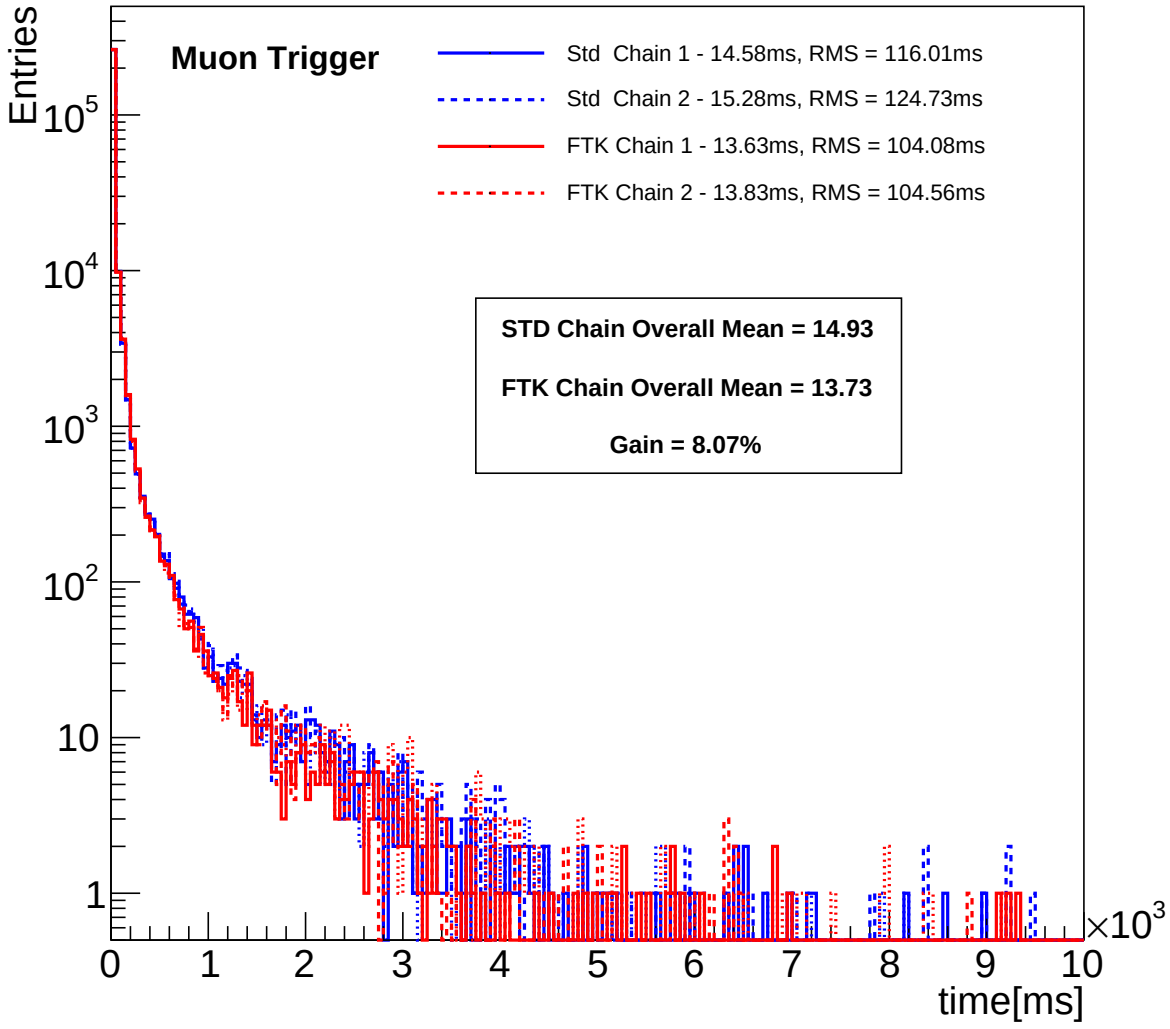


Figure 4.19: The CPU processing time per event in enhanced bias data, for both the standard chain (blue) and FTK chain (red). The dashed and dotted lines denote two separate processings to average over fluctuations in CPU load.

Chapter 5

Object Definitions

We demand rigidly defined areas of doubt and uncertainty!

– Douglas Adams

The analyses presented in Chapters 7, 8 and 9 share many commonalities, particularly in the techniques used to reconstruct physics objects from the detector hits and the systematic uncertainties on these. Tracks and clusters, built from the raw detector hits, are subsequently combined in various ways to reconstruct single particles like electrons and muons, jets of hadronic particles, and the missing transverse energy from particles that escape detection. This chapter describes these objects, with any differences between the analyses explicitly described.

5.1 Low Level Objects

Before attempting to reconstruct specific types of particle with the full ATLAS detector, low-level objects are reconstructed using the various sub-systems. In the calorimeters, *clusters* and *towers* are built, which group the energy deposits in neighbouring cells. In addition, *tracks* are built, separately for both ID and muon systems, which chart the passing of charged particles through these regions of the detector.

5.1.1 Clusters and Towers

The energy deposits in the calorimeters are grouped together into *clusters* or *towers* [125]. Towers are reconstructed using the *sliding window* algorithm. A calorimeter tower is the sum of energy in each layer of the calorimeter in an element of size $\Delta\eta \times \Delta\phi$. Towers are built in the region $|\eta| < 2.5$, and elements of size $\Delta\eta = \Delta\phi = 0.025$ are used, giving a total of 200×256 elements in $\eta - \phi$ space. A fixed window size of $N_\eta - N_\phi$ is then used to find the cluster from these towers.

In contrast, *clusters* are of variable size. Clusters are formed around seed cells which have an energy significance threshold $t = E_{\text{cell}}/\sigma_{\text{noise}} > 4$, where E_{cell} is the energy in that calorimeter

cell. The expected noise in the cell, σ_{noise}, originates from the read-out electronics and pileup, and is estimated from MC. Neighbouring cells, defined as the 8 cells surrounding the seed in the same layer as well as any cells overlapping the seed in the adjacent layers, are then added to these seed cells if they have $t > 0$. If $t > 2$, the procedure continues for the next set of neighbouring cells. Clusters with multiple local maxima of $E > 500$ MeV are split by repeating the clustering around these maxima and splitting shared cells by weighting the energy added to each cluster according to the local maxima energy and distance between them.

Clusters are calibrated to one of two energy scales. An *electromagnetic* (EM) scale cluster has been measured to the scale of an electromagnetically interacting particle, such as an electron [126], and this is used in the cluster building. A further calibration can be applied to these clusters, called the *local cluster weighting* (LCW) method, which improves resolution by classifying the cluster as having either an electromagnetic or hadronic origin, based on energy density and shower shape in the calorimeters. Hadronic clusters then have corrections applied, as derived from single charged and neutral pion simulations [127]. Finally, the four-vectors of all clusters are defined to have zero mass.

No systematic uncertainties are derived on clusters for the analyses in this thesis, as systematic uncertainties are defined directly on the high-level objects they build.

5.1.2 Tracks

The trajectories of charged particles passing through the ATLAS detector are reconstructed by building *tracks* in the inner detector and muon system. The magnetic fields from the ATLAS solenoid and toroids cause the paths forged by these particles to bend, and the amount of this bending can be used to infer particle properties: the bending direction indicates the charge, and the radius of curvature is inversely proportional to particle momentum. ATLAS contains two subdetectors for measuring tracks, with the ID able to measure all charged particles at the centre of the detector and the muon systems designed specifically to measure muon tracks in the layers furthest from the centre.

Inner Detector Tracks

The first step in building tracks in the ID is to group pixels and strips and with significant energy deposits. This is done with a connected component analysis [128], resulting in a collection of three-dimensional *space-points* which represent locations traversed by a charged particle. A pixel space-point is a single cluster of hits, while an SCT space-point must have clusters on both sides of the strip. An initial track building step is performed beginning with *track seeds* formed of three space-points. A Kalman filter [129] is used to build candidate tracks by requiring that at least one further space-point is compatible with the track seed, with multiple track candidates permitted per seed if there exists more than one space-point that is compatible. Candidate tracks

must subsequently pass the following quality criteria:

- $p_T > 400$ MeV
- $|\eta| < 2.5$
- A minimum of 7 pixel and SCT clusters
- A maximum of one shared space-point
- A maximum of two active but unmatched detector elements (*holes*), where a maximum of one hole is in the pixel detector
- The transverse impact parameter $|d_0^{\mathrm{BL}}| < 2.0$ mm [1]
- The longitudinal impact parameter $|z_0^{\mathrm{BL}} \sin\theta| < 3.0$ mm [2]

Remaining ambiguities resulting from these track candidates are resolved using a neural network trained to identify merged clusters that contain hits from multiple charged particles [130]. Those clusters identified as merged by the neural network are allowed to be used by multiple track candidates. Those that are not are identified as merged, but which are shared between multiple tracks, are kept in the best fit track and removed from the others. The final step is a tighter Kalman filter track fitting using all available information [131].

Uncertainties are derived directly on ID tracks for when these are used outside of the high-level objects [132], such as for defining uncertainties on jet substructure (see Sections 5.3.4 and 9.2). The uncertainty on track reconstruction efficiency is dominated by the material distribution of the ID, with smaller effects from MC modelling also relevant, leading to an overall uncertainty of around 0.5% at low $|\eta|$ up to around 2.5% at high $|\eta|$. The rate of fake tracks is estimated from MC, and an additional global factor of $\pm 10\%$ is applied as a systematic on these. Uncertainties from biases due to misalignment of the ID are also included for the impact parameters and track sagitta, though these are typically negligible.

Primary and Secondary Vertices

ID tracks are also used to construct *vertices*, points at which multiple particles interact and produce divergent tracks [133]. The *primary vertex* is defined as that which has the highest square sum of p_T in the tracks per event, $\sqrt{\sum p_T^2}$, with all other vertices labelled *secondary*. The primary vertex is used to define the location of the hard scattering interaction, while the secondary vertices may originate from pile up interactions, multi-parton interactions, or decays of long lived particles such as B-hadrons.

[1] The transverse impact parameter is a measure of the closest approach to the beamline in the transverse plane.

[2] The longitudinal impact parameter is the distance in the beam direction z, from the point at which the transverse impact parameter is determined, with θ the polar angle

Muon Spectrometer Tracks

For the identification of muons, tracks are built from hits in the muon spectrometers [134]. In the MDT and CSC chambers, a Hough transform [135] allows straight line fits to the hits in each layer to build segments in the plane of bending due to the magnetic field. This is complemented with the RPC and TGC hits, which measure the coordinate orthogonal to the bending plane. Track candidates are seeded initially from the middle layer and subsequently extended to the outer and inner layers. A track candidate must be built with at least two segments, with the exception of the transition region between barrel and endcap where a single high quality segment is sufficient. If two tracks share hits in the first two layers, but contain segments in the outermost layer none of which are shared, then both tracks are kept. This is done to preserve efficiency for close-by muons. Other segments which are shared between multiple tracks are assigned to the best fit track using a global χ^2 fit. Systematic uncertainties on MS tracks are built as part of the muon reconstruction described in Section 5.2.2.

5.2 Leptons

Throughout this document, the term *lepton* usually refers to electrons and muons. Although techniques exist to reconstruct τ leptons with ATLAS, they generally require dedicated analyses and thus are not extensively utilised here, though a veto is placed on events containing them in Chapter 8 with negligible effect on the results. Neutrinos can not be detected by ATLAS, though missing transverse momentum, discussed in Section 5.5, can often be used as a proxy. The following sections describe the methods used to reconstruct electrons and muons, along with the related systematic uncertainties.

5.2.1 Electrons

Electrons (and equivalently positrons) are reconstructed in the ATLAS detector by pairing energy in the electromagnetic calorimeter to tracks found in the inner detector. Electron clusters use calorimeter towers, initially using a sliding window of size 3×5 towers in $\eta - \phi$ space, with an efficiency of $\sim$ 100% for electrons above $E_T > 20$ GeV. These are subsequently paired to an ID track by matching in $\eta - \phi$ space. With a track matched, a larger sliding window of 3×7 (5×5) is subsequently used in the barrel (endcap) region to form the final electron cluster to which the track is matched [136]. These candidates are required to have $p_T > 25$ (27) GeV in the analysis presented in Chapter 7 (8 & 9) as well as $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$ in all cases, corresponding to the electromagnetic calorimeter acceptance after removing the crack region between the barrel and endcap. The higher p_T requirement in later years is due to the higher instantaneous luminosity in the LHC during these years, which meant that trigger thresholds had to be increased as discussed in Section 4.2.5. Further requirements are made on

the transverse impact parameter significance [3] $|d_0^{BL}\text{sig}| < 5$ and longitudinal impact parameter $|z_0^{BL}\sin\theta| < 0.5$ mm to mitigate pileup. After these selections the object is considered an "electron candidate". In order to discriminate between real electrons and other objects with similar signatures like photons or small jets, a multivariate likelihood ("LH") method is used, using inputs from each calorimeter layer (such as ratios of energy deposits and cluster widths), tracks (such as as number of hits or track impact parameter significance), and track-cluster matching (such as $\Delta\eta$, $\Delta\phi$ and ratio of cluster and track energies) to calculate a probability that the candidate is a real electron. Three working points are defined, referred to as "Loose", "Medium" and "Tight", depending on how strict the requirement is made on the LH output. The exact efficiency is dependent upon p_T and η, with the three working points corresponding to a selection efficiency of 93%, 88% and 80% respectively for an electron of $p_T = 40$ GeV [136]. The analysis presented in Chapter 7 uses tight electrons, while those in Chapters 8 and 9 use the loose working point. The electrons must further be "isolated", such that there is low activity in the detector around the identified electron. This requirement helps to select prompt[4] electrons by applying a threshold on the sum of energy in a cone of $\Delta R = 0.2$ around the calorimeter cluster or track. Two isolation requirements are used: the analysis in Chapter 7 utilises "Gradient" isolation, where efficiency is 90 (99)% at $p_T = 25$ (60) GeV, while Chapters 8 and 9 both utilise the "Fixed Loose" isolation requirements that $E_T^{\text{cone}}/p_T^{\text{electron}} < 0.20\,(0.15)$ for calorimeter cluster (track). Full details on electron identification and isolation algorithms in ATLAS can be found in [136].

The efficiencies of these algorithms are measured using the Tag and Probe method [124], with a data sample enriched in $Z \to ee$ and $J/\psi \to ee$ events. In this method, one electron produced in the heavy particle decay is reconstructed ("tagged"), leaving the other as an unbiased sample of "probes" that can be used to measure efficiency via a fit to the well known mass peak. In order to correct for differences in electron reconstruction algorithm performance in data and MC, the ratio of efficiencies measured in each case are used to derive a scale factor that is applied as an event weight to MC events, such that data and MC reaches good agreement. The values of these efficiencies are typically very close to unity in the case of electrons [136].

There are several sources of systematic uncertainties on electron reconstruction. Uncertainties due to reconstruction efficiency, including isolation and identification, are estimated by comparing results using alternative tag/probe identification requirements and background subtraction techniques, taking the mean result as the central value, and the root-mean-square (RMS) of these as the systematic uncertainty. The momentum scale and resolution of electrons is estimated in data and MC, with the MC subsequently corrected to the data, again using known

[3] The $|d_0^{\text{BL}}|$ significance, $d_0^{BL}\text{sig} = d_0/\sigma(d_0)$ where $\sigma(d_0)$ is the uncertainty in d_0^{BL}, is used to avoid discarding events with large uncertainty that might still pass subsequent ID requirements.

[4] A "prompt" electron (or muon) is defined as one produced from heavy particle decays (such as W/Z bosons). "Non-prompt" electrons or muons, which may originate from a photon conversion or heavy hadron decay, are usually considered backgrounds in lepton identification.

resonances such as $Z \to ee$ [137]. Systematic uncertainties related to pile up, calorimeter material and calibration, and MC modelling, are propagated as uncertainties on the electrons. In the analyses presented here, the systematic uncertainties due to electrons are typically very small and have minimal effect on the final results.

5.2.2 Muons

Muons are reconstructed by pairing an inner detector track with a muon spectrometer track via a global refit using hits from both subsystems. Hits in the MS may be added or removed to improve fit quality during this procedure. These combined tracks form muon candidates that must satisfy $p_T > 25$ (27) GeV for the analysis in Chapter 7 (Chapters 8 and 9), $|\eta| < 2.5$, and pass $|d_0^{BL}\text{sig}| < 3$ and $|z_0^{BL} \sin\theta| < 0.5$ mm. Further quality requirements are made on the number of hits in the ID: there must be ≥1 in the Pixel and ≥5 in the SCT, as well as ≤3 "holes" where no hits are present in a sensor covered by the track. Subsequently, muons are categorised into working points dependent upon the quality of the MS tracks. For the analyses in Chapters 8 and 9, the "medium" working point is used, which requires hits in ≥ 3 layers of the MDT, with the exception of those in the $|\eta| < 0.1$ region where the requirement is ≥ 1 layer with ≤ 1 hole. An additional requirement is made that $q/p_{\text{sig}} < 7$, where q/p_{sig} is defined as the difference in ratio of charge and momentum for the ID and MS tracks divided by the quadrature sum of the track uncertainties. Chapter 7 selects only "tight" muons, a subset of the medium WP which further requires hits in ≥ 2 MS layers and that the χ^2 of the combined ID and MS tracks is < 8. Similar to electrons, muons must also be isolated, in both the ID and the calorimeter. In the ID, the variable $p_T^{\text{varcone30}}$ is used, defined as the scalar p_T sum for additional tracks ($p_T > 1$ GeV) in a cone of radius $\Delta R = \min(10\text{ GeV}/p_T^{\mu}, 0.3)$. This p_T dependence improves performance at high p_T while keeping a reasonable cone size of 0.3 at low p_T. The calorimeter isolation uses $E_T^{\text{topocone20}}$, which uses a fixed radius of $\Delta R = 0.2$ to sum the transverse energy around the muon from clusters after subtracting any contribution from the muon itself, as well as correcting for pileup using the techniques of [138]. In Chapter 7, the same requirements are made for muons as for electrons: that efficiency is 90 (99)% at $p_T = 25$ (60) GeV, after applying thresholds on $p_T^{\text{varcone30}}/p_T^{\mu}$ and $E_T^{\text{topocone20}}/p_T^{\mu}$. Similarly Chapters 8 and 9 use fixed thresholds, this time $p_T^{\text{varcone30}}/p_T^{\mu} < 0.15$ and $E_T^{\text{topocone20}}/p_T^{\mu} < 0.3$. The reconstruction and identification efficiency is measured for muons with $p_T > 25$ GeV using $Z \to \mu\mu$ events with a tag and probe procedure, and the difference between data and MC measurements is applied as an event weight to MC, in a manner closely analogous to the electron case [134].

The systematic uncertainty on the reconstruction efficiency is below 1% in the full p_T range, as is the uncertainty due to the isolation requirement, and both originate primarily from the background subtraction at each stage. The uncertainty on the muon momentum scale and resolution are also applied, again derived using $Z \to \mu\mu$ events and a fit to the Z boson mass

peak [134]. The relative resolution is found to be approximately 2.9% in the endcap region and approximately 2.3% in the barrel, with good agreement between data and simulation. As with electrons, systematic uncertainties from muons have little effect on the results from any of the analyses presented in this thesis.

5.3 Jets

As quarks and gluons are color charged particles, they undergo the process of hadronisation, as discussed in Section 2.1. The consequence of this from an experimental point of view is that these particles do not hit the detector as a single cluster or track like the leptons discussed above. Instead, the experimental signature of a quark or gluon is a collimated spray of hadrons, and thus they are reconstructed as *jets*.

Many algorithms exist for jet reconstruction. The "Snowmass Conditions" [139] set out a series of requirements that a modern algorithm must meet. These state that it must be simple to implement both experimentally and theoretically, as well as result in sensible theoretical predictions in perturbation theory and be insensitive to the hadronisation model used. The most important result of this is that a jet algorithm must be *infrared* and *collinear* safe. That is, the jet reconstruction must not be modified by infinitely soft or small angle radiation. The most common family of algorithms in use today obey these requirements and use sequential recombination, defined with

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p})\frac{\Delta R_{ij}^2}{R^2}, \tag{5.1}$$

$$d_{iB} = k_{T,i}^{2p}, \tag{5.2}$$

$$\Delta R_{ij}^2 = (y_i - y_j)^2 - (\phi_i - \phi_j)^2. \tag{5.3}$$

The notations $k_{T,i}$, y_i and ϕ_i are the transverse momentum, rapidity and azimuthal angle of the input constituent respectively. These input constituents are most commonly calorimeter clusters or tracks, though any combination of detected objects could be used. The clustering begins by identifying the inputs with the smallest distance measure d. Should this be d_{ij}, the distance between inputs i and j, the inputs i and j are combined into a single input, the combined input is added to the list, the original inputs i and j are removed from the list, and the sequence continues. Should this be d_{iB}, the distance between the input i and the beam B, then i becomes a jet, is removed from the list of inputs, and the sequence begins again. This continues until all the inputs have been clustered into jets.

The exact behaviour of these algorithms is then determined by the parameters p and R. R is a size parameter that effectively determines the width in y-ϕ of the final jets. The parameter p determines how the ordering of the cluster sequence behaves. A positive value will cluster the constituents from softest to hardest, and the case $p = 1$ is commonly known as the k_T

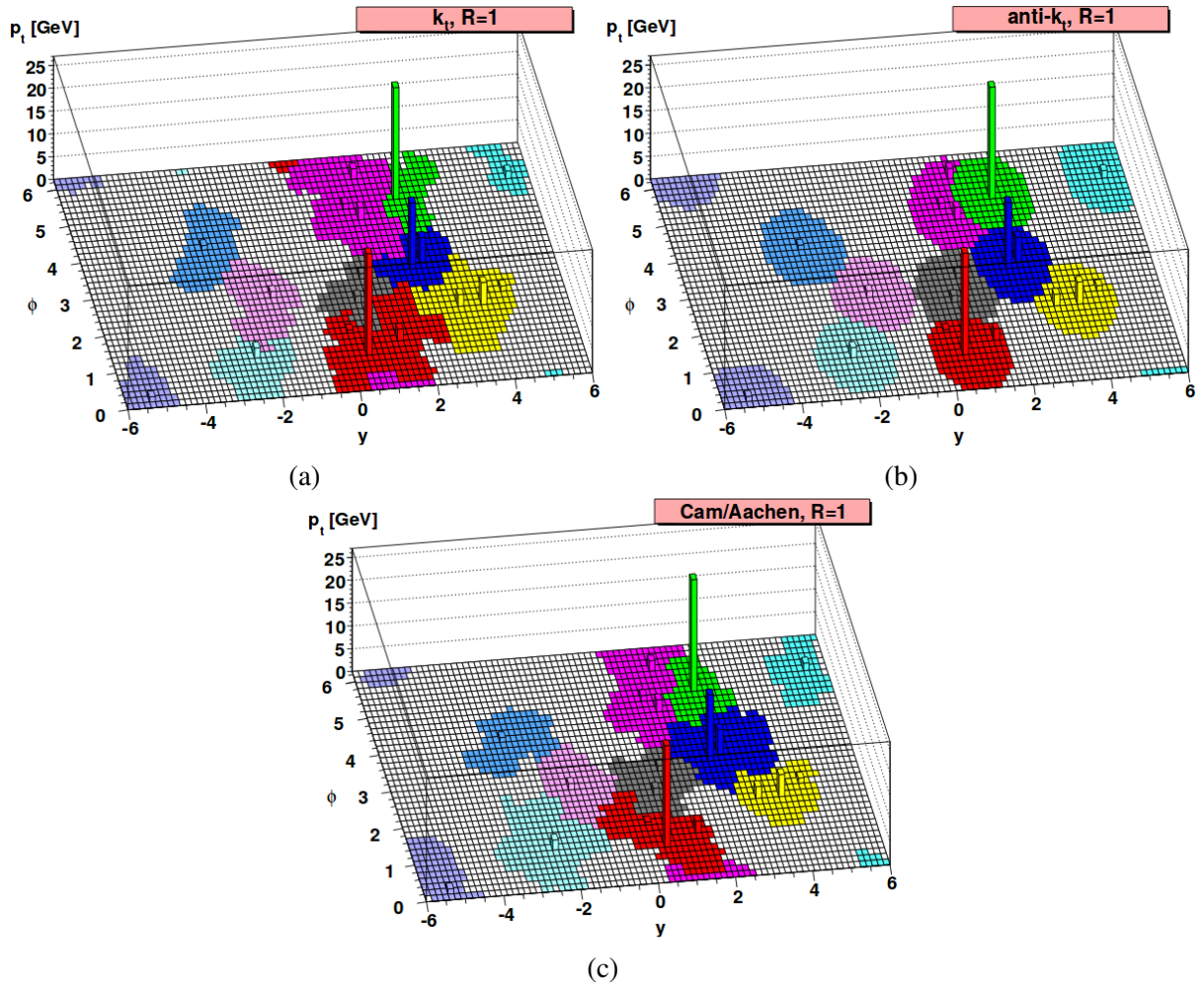


Figure 5.1: Jet finding results from an example event of the various jet algorithms described, for (a) k_T, (b) anti-k_T, and (c) Cambridge-Aachen.

algorithm [140]. $p = 0$ removes any dependence on the energy of the inputs, and is called the *Cambridge/Aachen* algorithm [141]. Negative values will begin clustering around the hardest input, and $p = -1$ is known as the *anti-k_T* algorithm [142]. Anti-k_T has the attractive property of creating approximately conic shaped jets around the hardest constituents (as shown in Figure 5.1) and is thus the most common algorithm in use today.

Several different definitions of jets are used in the analyses performed here, using various radius parameters and input constituents. All are reconstructed using the anti-k_T algorithm as implemented in the the FastJet package [143].

5.3.1 Small Jets

Jets reconstructed from EM-scale clusters with radius parameter R=0.4, referred to as *small jets*, are typically used to capture the hadrons originating from a single quark[5] or gluon. They

[5]Including *b*-quarks, but excluding top quarks; see Section 5.3.6.

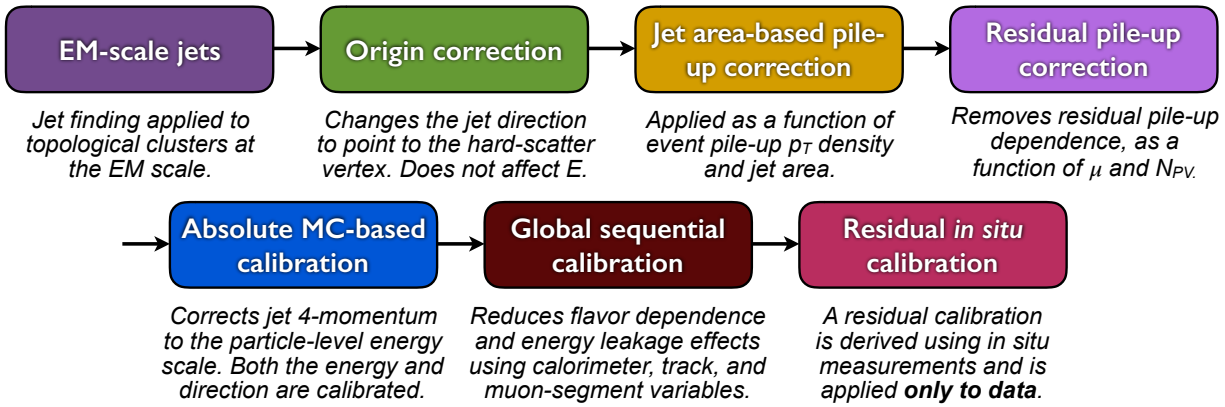


Figure 5.2: Cartoon summary of calibration procedure for small jets [145].

are reconstructed requiring $p_T > 25$ GeV and $|\eta| < 2.5$. In addition, a requirement on the multivariate quantity, the "jet vertex tagger" (JVT), is made for jets with $p_T < 50$ GeV and $|\eta| < 2.4$, to reduce pile up effects. JVT takes into account 2 track based variables to perform a 2 dimensional likelihood fit to determine the probability for a jet to be the result of pileup. Full details are given in [144].

Small jets are calibrated in several stages, summarised in Figure 5.2. After initial clustering, all jet axes are aligned to the centre of the detector, so the first step is to *origin correct* such that the jet axes points to the primary vertex. Subsequently, corrections are applied to account for pile up effects, using fits to the jet p_T and η. Three factors are used to determine the correction: the fraction of p_T in each jet that is expected to originate from pile up, based on the jet area; the number of primary vertices in the event; and the average number of pp collisions per bunch crossing.

The next step in the small jet calibration is an MC based calibration of the jet energy scale (JES) and jet η. Reconstructed jets are matched to truth jets via a $\Delta R < 0.3$ requirement, only if there is no other reconstructed jet within $\Delta R = 0.6$ nor any other truth jet within $\Delta R = 1.0$, leaving only well matched and isolated jets. The mean energy response is calculated from these jets via a Gaussian fit to the $E^{\text{reco}}/E^{\text{truth}}$ distribution, binned in E^{truth} and η_{det}, the jet η relative to the detector centre, which avoids ambiguity on which part of the detector the jet lies in and thus clearly separates effects due to different detector technologies. This is derived separately for full and fast ATLAS simulation, as discussed in Section 6.7, with an additional non-closure uncertainty applied to fast simulation samples due to the approximations made modelling the calorimeter.

A *global sequential calibration* (GSC) is the next stage of jet calibration. After the MC based calibration, some residual dependence is observed on a number of observables related to longitudinal and transverse features of the jet, such as the particle composition and shower shape. The GSC procedure independently corrects, as a function of $p_{T,\text{truth}}$ (E_{truth}) and $|\eta_{\text{det}}|$ for the first four (fifth) observables, in a sequential, independent manner similar to the MC

calibration described above. The observables corrected for are:

- f_{Tile0} : jet energy fraction in the first layer of the Tile
- f_{LAr3} : jet energy fraction in the third layer of the LAr
- n_{track} : number of tracks ghost associated [146] to the jet with $p_T > 1$ GeV
- $\mathcal{W}_{\text{track}}$: average & p_T weighted transverse distance in $\eta - \phi$ between jet axis and ghost associated tracks
- n_{segments} : number of track segments in the Muon Systems ghost associated to the jet (*punch-through*)

Including correlations between observables, or changing the order of the sequence, does not improve the resolution of the jets [145]. The average jet energy is kept constant for each of the five corrections, meaning only the dependence on the observables is corrected for in this procedure, with the overall JES conserved.

The final steps in the small jet calibration are in situ calibrations comparing data and MC. Firstly, an η-intercalibration, which uses jets from the well measured central region $|\eta_{\text{det}}| < 0.8$ to calibrate jets in the forward region $0.8 < |\eta_{\text{det}}| < 4.5$, using dijet events back to back in ϕ, is performed. Three subsequent calibrations are then carried out for jet p_T, where the balance of central jets and well known physics objects is used to measure the p_T response. First, two studies, utilising photons and leptonically decaying Z bosons, calibrate the low to mid p_T range. These are performed in parallel and combined, before finally, several low p_T jets are balanced against a single high p_T jet, to complete the calibration across a full and complementary p_T range from 20 GeV to 2 TeV. There is good agreement found between methods, as shown in Figure 5.3. A large number of systematic uncertainties on the JES, coming from each of these in situ methods, are propagated through the full procedure. This includes uncertainties on the selection and measurement of objects used for the balance, the quality of the MC simulation, finite MC statistics, and other assumptions used in the event topologies. Additionally, the jet energy resolution (JER) is estimated from the balance distributions in the in situ calibrations [147].

The total uncertainty on the jet energy scale is shown in Figure 5.4. The largest contributions come from the in situ calibrations. The next largest contributions are flavor response and composition, the difference in response between quark and gluon jets with relative compositions taken from MC. The full breakdown of uncertainties can be found in [145].

5.3.2 B-Tagging

Small jets originating from b-quarks are phenomenologically distinct from those originating from light quarks, due to the presence of B-hadrons. The long lifetime of these leads to the

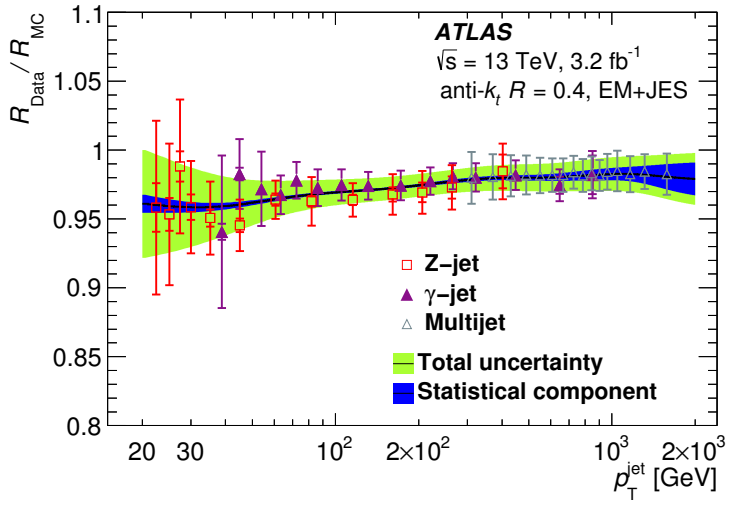
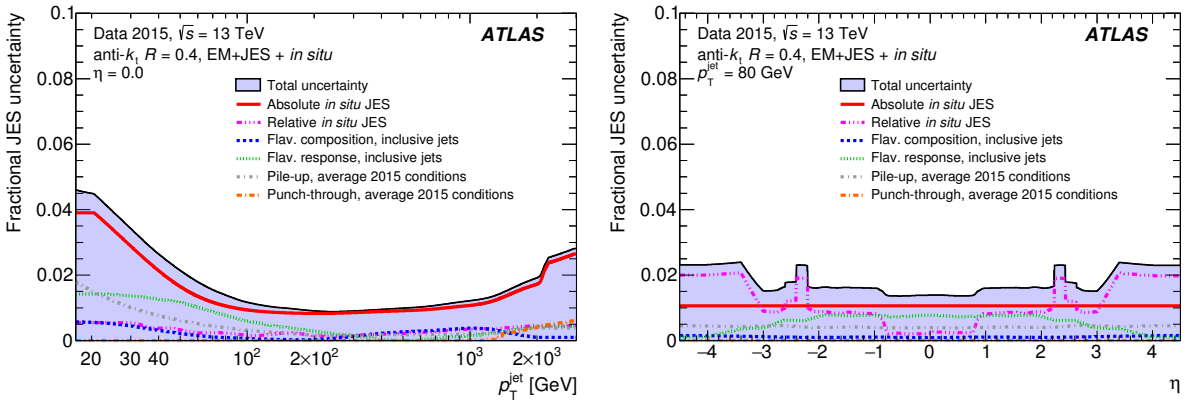


Figure 5.3: The results of the in situ calibrations as a function of jet p_T, showing the complimentarity of Z-jet (red squares), γ-jet (filled purple triangles), and multijet (black unfilled triangles) methods. The y axis shows the double ratio $R_{\text{data}}/R_{\text{MC}}$, with R the ratio of the jet and the balancing object(s) in each method [145].



(a) (b)

Figure 5.4: The full uncertainty on the JES for (a) a jet with $|\eta| = 0$, as a function of p_T, and (b) a jet with $p_T = 80$ GeV, as a function of η [145]. Each line represents a different sub-component of the uncertainty, with the total shown represented by the shaded blue area.

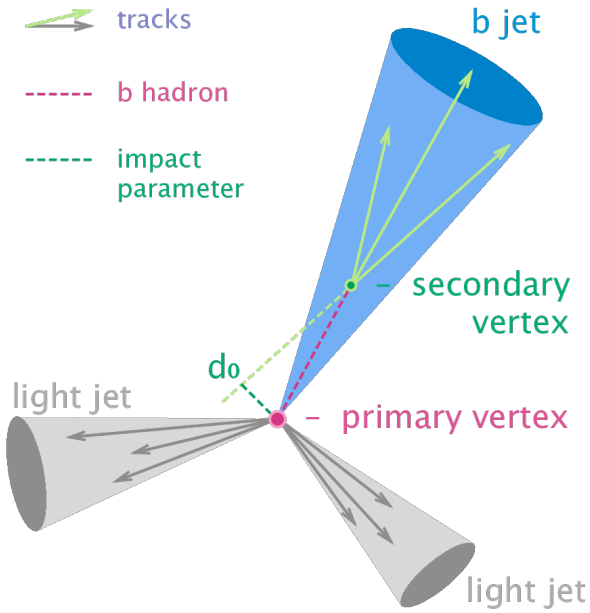


Figure 5.5: Diagram demonstrating the characteristic displaced vertex in a small jet originating from a B-hadron [151]. This distinctive property can be used to b-tag these jets, to distinguish them from small jets originating from charm quarks, light quarks, or gluons.

presence of secondary vertices and displaced tracks, as shown in Figure 5.5. These properties can be used to *b-tag* the jet. This is done with a multivariate discriminant known as the MV2 algorithm [148–150] trained to discriminate between jets originating from bottom, charm, and *light* particles, where light is inclusive of u, d, and s quarks as well as gluons. Based on a Boosted Decision Tree, MV2 combines information from three complementary varieties of b-tagging algorithms. All of these algorithms utilise tracks, which are matched to the calorimeter jets via a p_T dependent ΔR requirement that results in narrower cones for higher p_T jets with more collimated decay products. In this section, the term "b-jet" refers to a jet truth matched to a B-hadron, while "b-tagged" refers to a jet passing the MV2 algorithm selection.

Impact Parameter Algorithms: IP2D & IP3D

The first of the algorithms combined in MV2 utilise information on the impact parameters of the tracks associated to the jet. The two dimensional algorithm, IP2D, includes only the transverse impact parameters, while the three dimensional algorithm, IP3D, also incorporates the longitudinal impact parameters. Both IP2D and IP3D are utilised, as although IP3D contains more information, it is also more sensitive to contamination from pile up due to the longitudinal impact parameter typically being larger for pile up jets. Log-likelihood ratios are constructed for each combination of b, c, and light quark hypotheses, giving six total variables that are used as input to the MV2 BDT [148].

Secondary Vertex Finding: SV

The second algorithm used as input to MV2 is a vertex reconstruction algorithm. By reconstructing a secondary vertex from the tracks associated to the jet, the properties of this vertex can be used to discriminate between b, c, and light jets. In total, 8 variables calculated from the secondary vertex are included in the BDT training, including the invariant mass of tracks matched to the vertex, the ΔR between jet and SV, the jet energy fraction, the number of associated tracks, and transverse and 3D distance measures [148].

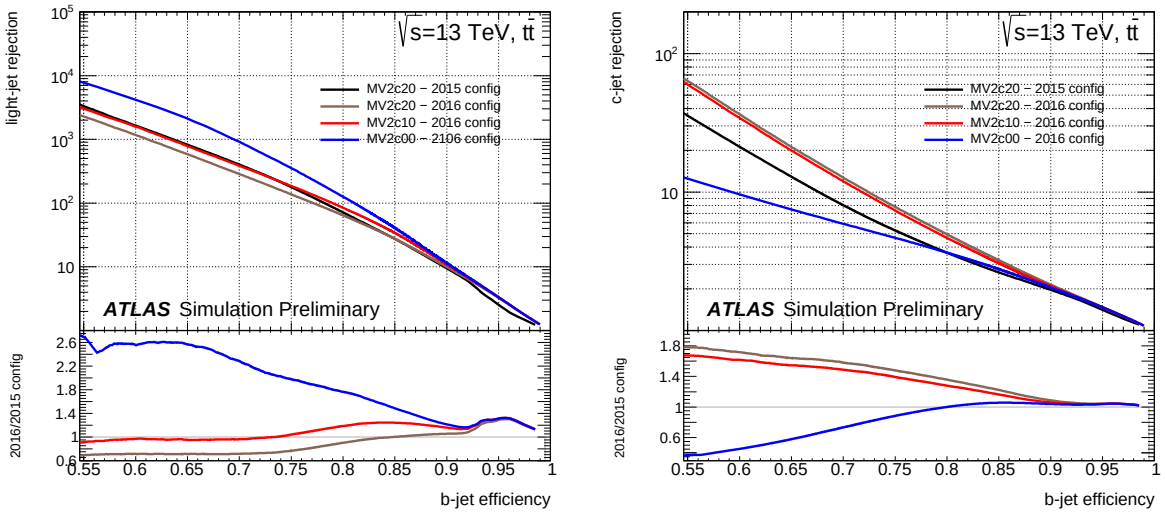
JetFitter

JetFitter is a Kalman Filter [129] based algorithm that reconstructs the entire decay chain from PV through B- and C-hadron decays. This provides similar but complementary information to the SV finding by reconstructing the full path of the decay, allowing multiple secondary vertices including some with a single track. JetFitter provides a similar 8 variable list to the BDT as the SV algorithm, this time based on the full path information [148].

Performance of the MV2 Algorithm

Along with the inputs from the three algorithms above, the jet p_T and η are given to the BDT, in order to exploit correlations with the other discriminating variables. Since these kinematics should not directly discriminate between signal and background, the signal distribution is reweighted to the background distribution to prevent this. The training is then performed with b-jets as signal, with a combined sample of charm and light jets constituting the background. The relative fraction of charm and light jets used in the training influences the relative rejection performance of these two backgrounds, and so this fraction can be tuned depending on performance priorities. Four *working points* (WP) are then defined by placing a threshold on the BDT output that achieves a set b-tagging efficiency in an MC sample independent of the training sample. These working points correspond to 60%, 70%, 77%, and 85% b-tagging efficiency. Each of these working points are then independently calibrated in data as described below, and scale factors are applied to correct the MC to the data. In addition, one further scale factor is derived, for a *pseudo-continuous* calibration. Here, the BDT output is separated into five bins, corresponding to each working point and bounded at 0% and 100%, and calibrated together. This allows for information from each working point to be used together in an analysis, instead of the more traditional binary tagging approach using a single fixed working point.

The analysis in Chapter 7 uses an admixture of 80% light and 20% charm jets as background in the BDT training, and this training is referred to as the MV2C20 algorithm [148] and is optimised for 2015 data taking conditions. The analyses in Chapters 8 and 9 use 93% light and 7% charm in the training, referred to as the MV2C10 algorithm [149]. These fractions were chosen after it was found that, while re-optimising training for 2016 data, the light jet rejection



(a) (b)

Figure 5.6: The rejection rates for (a) light jets, and (b) charm jets, for different configurations of the MV2 algorithm [149]. The black line corresponds to the performance in Chapter 7, while the red line corresponds to the performance in Chapter 8.

improved significantly with respect to the MV2c20 method, while charm jet rejection was only slightly degraded. This is shown in Figure 5.6.

Calibration

The calibration of b-tagging efficiency is performed with a maximum likelihood fit, which can take into account the per-event jet correlations by determining f_1 and f_2, the fractions of events with one and two b-tags respectively. Dileptonic $t\bar{t}$ events [150] are selected, requiring two oppositely charged leptons and at least two small jets, at least one of which is b-tagged at the 85% WP. Events are further selected using a BDT trained to separate events containing two b-jets from events with fewer than two b-jets. Only kinematic information that is minimally correlated to the flavor of the jets is used in the training to prevent introducing any bias in the MV2c10 distribution. The end result is a selection with a high purity of b-jets. The likelihood is constructed per event using probability density functions determined in MC with the p_T and MV2 score of each jet, with the b-tagging efficiency extracted from data. The resulting scale factors are binned in p_T and are typically of the order of 1-3%. The largest systematic uncertainties on the results are those related to the MC modelling, with additional uncertainties from the detector objects, background cross-sections, and extrapolation for higher p_T jets.

As well as the actual b-tagging efficiency, the mistag rates of charm and light jets must also be calibrated. The charm calibration is performed in two different topologies, one in $t\bar{t}$ [152] and one in $W + c$ [153] production, and subsequently combined. The light jet calibration is also performed using two complementary methods, which produce results in good agreement [154]. The resulting b, c, and light scale factors from these calibrations are applied event-by-event by

multiplying together the per-jet scale factors.

Due to the complex determination of the full b-tagging calibration, the number of components of the systematic uncertainties is very large. To reduce this, an eigenvalue decomposition of the covariance matrix is performed, following the procedure described in [150]. The exact number of resulting eigenvalues differs per analysis, depending on correlation scheme chosen, with the trade off of the number of components against how much correlation information is retained. The analysis in Chapter 7 uses a total of 12 eigenvalues, while the analysis in Chapter 8 considers a total of 110, reflecting the more complex strategy of pseudo-continuous tagging and greater sensitivity to b-tagging of the $t\bar{t}H, H \to b\bar{b}$ analysis. Section 9.3 uses an updated calibration with respect to Chapter 8, which retains 80 components in total.

5.3.3 Track Jets

In Section 9.1, a study is performed of an alternative event selection utilising *track jets* for b-tagging, instead of small jets. Track jets are built and calibrated with the same methods as for small jets, but instead of using calorimeter clusters as constituents, they use tracks from the inner detector. Since the ID has improved angular resolution compared to the calorimeter, track jets built with a radius parameter R=0.2 can be used, which may suit boosted analyses due to the increasing collimation of decay products at high p_T. For example, two close-by b-quarks might be identifiable as two distinct R=0.2 jets, instead of a single R=0.4 jet. Since tracks only measure charged particle activity, these jets have a lower energy scale than calorimeter jets, and thus the requirement for a track jet is $p_T > 10$ GeV. These track jets, along with R=0.4 track jets also used for cross checks in Section 9.1, utilise the same MV2C10 b-tagging algorithm as described in Section 5.3.2.

5.3.4 Large Jets

A *large jet* is defined as an anti-k_T R=1.0 jet with cluster inputs, calibrated to the LCW scale as described in Section 5.1.1. Jets of this size are typically used to reconstruct boosted heavy objects, such as top quarks, Higgs bosons, and W/Z bosons, as described in Section 3.1. They are required to have $p_T > 200$ GeV and $|\eta| < 2.0$. The tighter $|\eta|$ requirement compared to small jets ensures the large jet is fully contained in the inner detector. The large jets are then *trimmed* [155] in order to reduce the impact of soft radiation and pileup. In this procedure, subjets of radius R_{sub} are formed inside the large jet, and the constituent clusters of these subjets are removed from the large jet if the fraction of p_T in the subjet is smaller than a set threshold, $\frac{p_{T,sub}}{p_{T,large}} < f_{cut}$. The parameters used are $R_{sub} = 0.2$ and $f_{cut} = 0.05$, optimised for the 2015 data taking period. In this way, soft clusters originating from secondary sources such as pile up, underlying event, or final state radiation, are removed from the jet, leaving just the hard deposits that are likely to originate from a heavy particle decay, such as a top quark or Higgs boson.

The JES calibration of the large jets is performed using a MC based technique, similar to that of [156] and modified as in [157], using a Pythia [158] MC sample that includes the effects of pile up. Firstly, the jet axis is origin corrected in the same way as the small jets described in Section 5.3.1. Subsequently, the jet energy is corrected using the response calculated from truth MC, and further η is corrected to account for bias in the instrumental precision in various regions of the detector. The magnitude of these corrections are typically $\mathcal{O}(1\%)$ and $\mathcal{O}(0.01)\%$ respectively.

Uncertainties on large jets are calculated using the R_{track} procedure [157]. This is an in situ, data driven approach that takes advantage of the two independent measurements in the ID and calorimeter systems. A track jet with R=1.0 is built in the same region of the detector as the calorimeter jet, and then the ratio r^x of a given variable x is computed as:

$$r^x = \frac{x_{\text{calo}}}{x_{\text{track}}}. \tag{5.4}$$

The double ratio, computed using r^x as measured in data and MC, can then be defined as:

$$R^x = \frac{r^x_{\text{data}}}{r^x_{\text{MC}}}. \tag{5.5}$$

This double ratio can then be used to understand how well a given quantity x can be measured. The data/MC comparison gives one component of the systematic uncertainty, comparison of various MC generators gives a second, and propagation of tracking uncertainties (see Section 5.1.2) gives a third. Each of these components are considered fully correlated for p_T and mass. These systematic uncertainties are much larger than the uncertainties on other detector level objects in the analyses presented here, and are dominant in the analysis presented in Chapter 7. Finally, the uncertainty on the jet mass resolution (JMR) is obtained by artificially degrading the JMR by 20% via smearing. This uncertainty is typically very small in comparison to the JES systematics.

5.3.5 Jet Substructure

When looking for high p_T heavy particles using large jets, it becomes important to be able to distinguish between those jets originating from different parent particles; the three-prong decay of a top quark is in many ways distinct from the two-body decay of a W or Z boson, for example. In recent years, a rich phenomenology of *jet substructure* (JSS) observables have been developed in order to probe these distinctions. An insightful pedagogical overview is given in [159]. Broadly speaking, there are two classes of JSS observables: those which are sensitive to the clustering history of the large jet, typically with units of mass; and those which are sensitive to the number of hard particles ("prongs") resulting from the heavy particle decay.

The most common of the first type of JSS observable is the *splitting scale*, which describes

the scales of the jet at specific points in the clustering procedure [160]:

$$d_{ij} = \min(p_{T,i}^2, p_{T,j}^2)\Delta R_{ij}^2 / R_0^2, \tag{5.6}$$

where $p_{T,i}$ refers to the jet after clustering into i proto-jets, R_0 is the jet radius parameter, and ΔRij is the angular distance between the jets at stage i and j. Thus, $\sqrt{d_{ij}}$ is an observable, with units of mass, which measures the step size in the clustering at the ith stage. Most commonly, one is interested in $\sqrt{d_{12}}$ or $\sqrt{d_{23}}$, the splitting scale at the final and the penultimate clustering stage respectively. Another common JSS variable to probe similar qualities is Q_W, defined as the minimum invariant mass of a pair of subjets formed when the large jet is decomposed into 3 exclusive k_T jets. This will capture information on other structure contained within only a part of the large jet, such as an intermediate W boson in a top quark decay.

The second class of JSS variable encapsulates the *N-subjettiness* (τ) series of observables [161] as well as the *energy correlation functions* (ECF) [162]. The τ series are defined as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k} ... \Delta R_{N,k}\}, \tag{5.7}$$

where the sum k runs over the constituents in the jet, each with transverse momentum $p_{T,k}$. The large jet is decomposed into N exclusive k_T subjets labelled $1...N$, and $\Delta R_{1,k}$ is the angular separation between the subjet 1 and particle k. The normalisation factor d_0 is defined as:

$$d_0 = \sum_k p_{T,k} R_0, \tag{5.8}$$

with R_0 the jet radius of the original large jet. Thus, the variable τ_N can be thought to represent how well a large radius jet can be reconstructed as N smaller jets, with low values implying radiation all aligned along the subjet axes and large values implying radiation away from these axes. In practice, it is usually most useful to consider ratios of these observables:

$$\tau_{ij} = \frac{\tau_i}{\tau_j}. \tag{5.9}$$

Thus, τ_{21} or τ_{32} are observables representing how well a large jet can be described by 2 or 3 subjets respectively. For example, a boosted top quark, with 3 hard prongs in its decay, would be expected to have a lower value of τ_{32} than a W boson or gluon-initiated jet.

Similar to the N-subjettiness observables, the ECFs measure the radiation pattern inside a large radius jet. The first three ECFs are defined as

$$e_1^\beta = \frac{1}{p_{T,J}} \sum_{i \leq J} p_{T_i}, \tag{5.10}$$

$$e_2^\beta = \frac{1}{p_{T,J}^2} \sum_{i \leq i < j \leq J} p_{T_i} p_{T_j} R_{ij}^\beta, \tag{5.11}$$

$$e_3^\beta = \frac{1}{p_{T,J}^3} \sum_{i \leq i < j < k \leq J} p_{T_i} p_{T_j} p_{T_k} R_{ij}^\beta R_{ik}^\beta R_{jk}^\beta, \tag{5.12}$$

with the sums running over the number of constituents J inside the large jet. β is an angular exponent, controlling the impact of the angular separation, which is most often chosen as $\beta = 1$. Dimensionless functional forms of the ECFs can be built to discriminate particular features of a jet. One particular example of this is the D_2 variable

$$D_2 = e_3 / e_2^3. \tag{5.13}$$

This variable is built in order to provide 2-prong discrimination, in a similar manner to τ_{21}. It will take small values for jets that are well described by the two prong hypothesis, and large values for jets that are not well described by this.

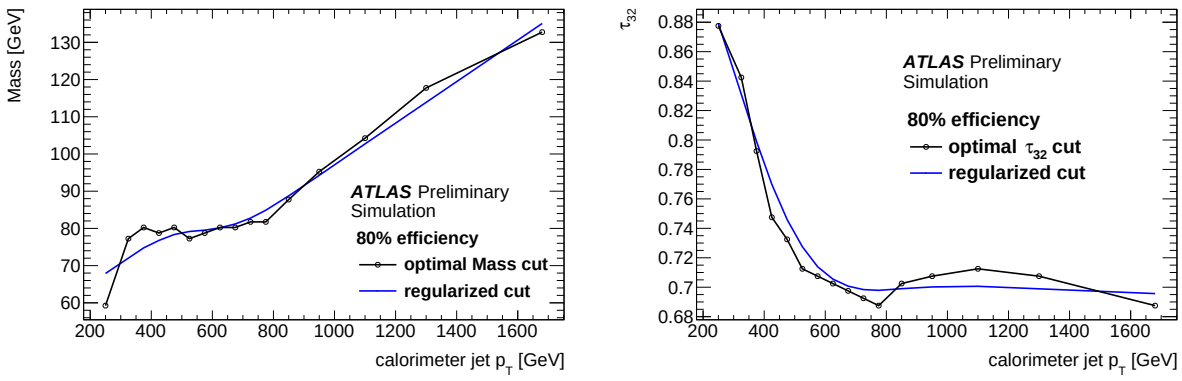
Further discussion of jet substructure, in the context of reclustered jets, can be found in Section 9.2.

5.3.6 Large Jet Top-tagging

The large jets described in Section 5.3.4 are further *top-tagged*, if they satisfy a set of requirements designed to identify them as originating from a boosted top quark. In Chapter 7 and Section 9.1, this is done using large jets with $p_T > 300$ GeV using the ATLAS Early Run-II top tagger [163]. This uses requirements on two substructure variables – the calibrated mass of the jet and the N-subjettiness ratio τ_{32}.

The exact value of the requirements on large jet mass and τ_{32} are p_T dependent in order to maintain an approximately flat efficiency in the full p_T range, as determined in MC. The required values for a fixed efficiency of 80% are evaluated explicitly in bins of p_T and then smoothly interpolated to avoid fluctuations. These range from $m > 70$ GeV and $\tau_{32} < 0.85$ for a jet of $p_T = 200$ GeV, to $m > 135$ GeV and $\tau_{32} < 0.0.7$ for a jet of $p_T \geq 1600$ GeV, with the precise values across the full range shown in Figure 5.7. The resulting efficiency, calculated in a $Z' \rightarrow t\bar{t}$ sample, in each bin of p_T is shown in Figure 5.8, showing an approximately flat efficiency, though performance drops off for $p_T < 350$ GeV. This is not surprising, since the rule of thumb $\Delta \simeq \frac{2m}{p_T}$ would imply that an $R = 1.0$ jet would not always fully contain a top quark (of mass $m \simeq 175$ GeV) decay until ~ 350 GeV.

The uncertainty on τ_{32} is derived using a similar R_{track} method as described in Section



(a) (b)

Figure 5.7: The requirements applied for the top-tagging algorithm as a function of jet p_T, for (a) mass and (b) τ_{32}. The black points show the empirically determined requirement for 80% efficiency, with the blue lines showing the smoothed interpolation [164].

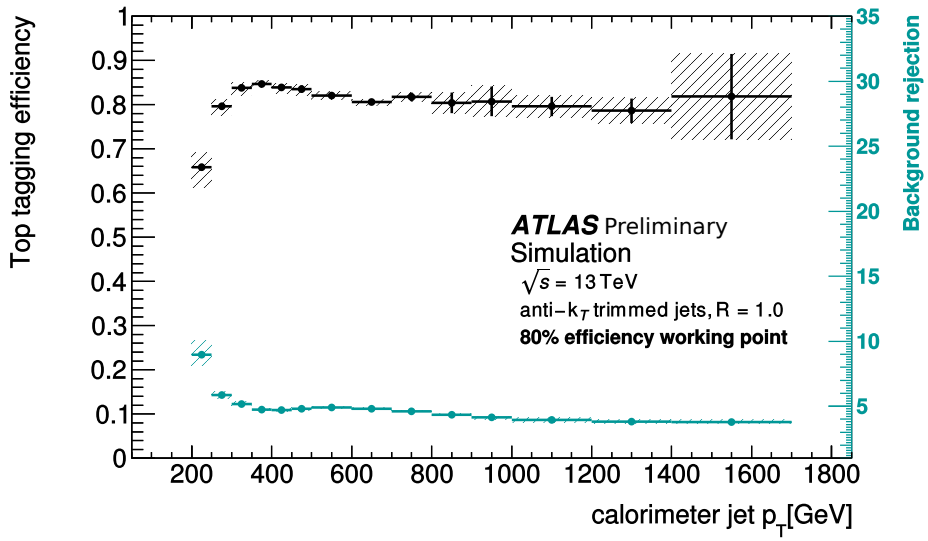


Figure 5.8: The tagging efficiency and background rejection for top quarks as a function of jet p_T. The efficiency is approximately flat above $p_T > 350$ GeV, with the drop off below this value due to the insufficient collimation of the decay products to be reliably captured by the R=1.0 jet [164].

5.3.4. It is considered uncorrelated with the p_T and mass uncertainties, since it is only weakly correlated with the overall JES and is more sensitive to the underlying jet radiation pattern. The uncertainty on this variable as derived in $\sqrt{s} = 8$ TeV is considered as one component, with a second cross-calibration component taking into account differences in LHC conditions, ATLAS upgrades, and reconstruction settings at $\sqrt{s} = 13$ TeV also applied separately.

5.3.7 Reclustered Jets

An alternative technique for reconstructing boosted objects, that does not suffer from the large systematic uncertainties associated with large jets built directly from clusters, is to use *reclustered* (RC) jets. In this method, the small jets described in Section 5.3.1 are used as input constituents to the anti-k_T algorithm, which reclusters these into R=1.0 jets. This allows direct propagation of the systematic uncertainties associated with the input small jets, which are already calibrated, with no need for any further calibration or uncertainties. It also means that b-tagging associated with large jets can be done directly on constituent subjets, instead of using ΔR matching, and combination with resolved channels is simplified as the same objects are used. The downside to this method is that it is not possible to define some substructure variables directly. The variables $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$, as defined in Section 5.3.5, can be calculated using the subjet clustering history, assuming > 2 or > 3 subjets respectively. Similarly, the Q_W variable can be calculated simply using the subjets directly. Other variables, such as τ_{21}, τ_{32}, and the energy correlation functions and associated ratio variables such as D_2, require the full cluster information and thus are not well defined using subjet inputs. A new method for defining jet substructure variables in reclustered jets is discussed in Section 9.2.

Reclustered jets are required to have at least two R=0.4 constituent subjets, $p_T > 200$ GeV, $|\eta| < 2.0$, and $m > 50$ GeV. Trimming is applied, using $R_{sub} = 0.2$ and $f_{cut} = 0.05$, although this has minimal effect since the constituent small jets have already had pile up suppression techniques applied, via the JVT requirement.

5.4 Overlap Removal

The electron, muon, and jet reconstruction algorithms are run independently, and a procedure is thus required to remove overlapping objects. The angular separation between two objects, ΔR, is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{5.14}$$

where $\Delta\eta$ and $\Delta\phi$ are the difference in pseudorapidity and azimuthal angle between the two objects in question. In order to prevent the same calorimeter deposits and tracks being used in identification of more than one object, overlap removal is performed as follows:

- If an electron shares a track with a muon, the electron is removed
- If a small jet is within $\Delta R < 0.2$ of an electron, the small jet is removed.
- If an electron is within $\Delta R < 0.4$ of a small jet (after the previous step is performed), the electron is removed.
- If a small jet is within $\Delta R < 0.4$ of a muon, and has less than three tracks, the small jet is removed. If it has three or more tracks, the muon is removed.

Large jets may also overlap with small jets or leptons, but this may be desired depending on the analysis details. For example, in Chapter 7, small jets overlapping the large jet (those with $\Delta R < 1.0$) are checked for b-tags as part of the event selection. Overlap with electrons or muons is removed by requiring that the large jet and the lepton are separated by $\Delta\phi(\Delta R) > 1.0$ in Chapter 7 (9.1).

5.5 Missing Transverse Momentum

In hadron collisions, it is impossible to know the fraction of the centre-of-mass energy contained in the colliding partons, and so the expected sum of the momenta in detected objects is not known. However, since the beams collide approximately head on in the z direction, the sum of the momenta transverse to the beam line would be expected to sum to zero if all objects were detected. If the momentum does not sum to zero, the difference from zero is referred to as the *missing transverse momentum* $\boldsymbol{p}_T^{\text{miss}}$. The magnitude of this vector is the *missing transverse energy* E_T^{miss}, with direction ϕ^{miss}. $\boldsymbol{p}_T^{\text{miss}}$ could be the result of a mismeasurement of detected particles, or of undetected particles. Since neutrinos can not be detected by the ATLAS experiment, $\boldsymbol{p}_T^{\text{miss}}$ is often used as a proxy measurement of neutrinos[6] in reconstructing events such as those including a leptonically decaying W boson.

$\boldsymbol{p}_T^{\text{miss}}$ is constructed by summing the x and y components of each reconstructed lepton, photon and small jet, as well as an additional term for soft radiation built from ID tracks that originate from the primary vertex but which are not associated with any other reconstructed object. To ensure that no energy contribution is double counted, $\boldsymbol{p}_T^{\text{miss}}$ is calculated after overlap removal. ID tracks are used for the soft term due to their excellent resolution in the z direction enabling good association with the primary vertex – neutral contributions from the calorimeter are not included due to the large contribution from pile up [165].

The uncertainty on $\boldsymbol{p}_T^{\text{miss}}$ is determined by propagating the scale and resolution uncertainties of all the objects used in the calculation. The analyses presented here do not rely heavily on $\boldsymbol{p}_T^{\text{miss}}$, and so these uncertainties are typically negligible.

[6] Of course, $\boldsymbol{p}_T^{\text{miss}}$could also be a result of BSM particles escaping the detector.

Chapter 6

Monte-Carlo Simulation

Reality is frequently inaccurate.

– Douglas Adams

In order to understand the data collected by the detectors, simulations are extensively utilised in high energy physics. Due to the probabilistic nature of quantum mechanics, *Monte-Carlo* (MC) techniques, based on random number generation, are used to sample from the probability distributions inherent in such calculations and build individual simulated particle collision events. However, the calculations become incredibly complex incredibly quickly, and in order to generate sufficient statistics, it is necessary to make many assumptions and approximations. To this end, these calculations are divided into a number of steps; matrix element generation, parton showering, hadronisation, and underlying event / multiparton interaction (UE/MPI). This breakdown is depicted in Figure 6.1, and discussed over the following sections, along with discussion of detector simulation in Section 6.7 and of the samples used in the presented analyses in Section 6.8.

6.1 Factorisation Theorem

An important concept in QCD is the *factorisation theorem*, which states that the hard scattering cross-section $\sigma(P_1, P_2)$ for two hadrons with momentum P_1 and P_2 can be factorised into separate terms for the short distance and long distance interactions [166], according to the equation

$$\sigma(P_1, P_2) = \sum_{ij} \int dx_1 \; dx_2 \; f_i(x_1, \mu_f^2) \; f_j(x_2, \mu_f^2) \; \hat{\sigma}_{ij}(p_1, p_2, \alpha_S(\mu_f^2), Q^2/\mu_f^2). \tag{6.1}$$

Here, the partons p_i involved in the hard scattering interaction have momentum $p_i = x_i P_i$. The scale of the interaction is denoted by Q. Thus, Equation 6.1 factorises into two pieces; one which determines the probability to have partons with momentum p_i, and a second piece which is the probability of the parton-parton interaction, and as a result the calculation of hadron-hadron

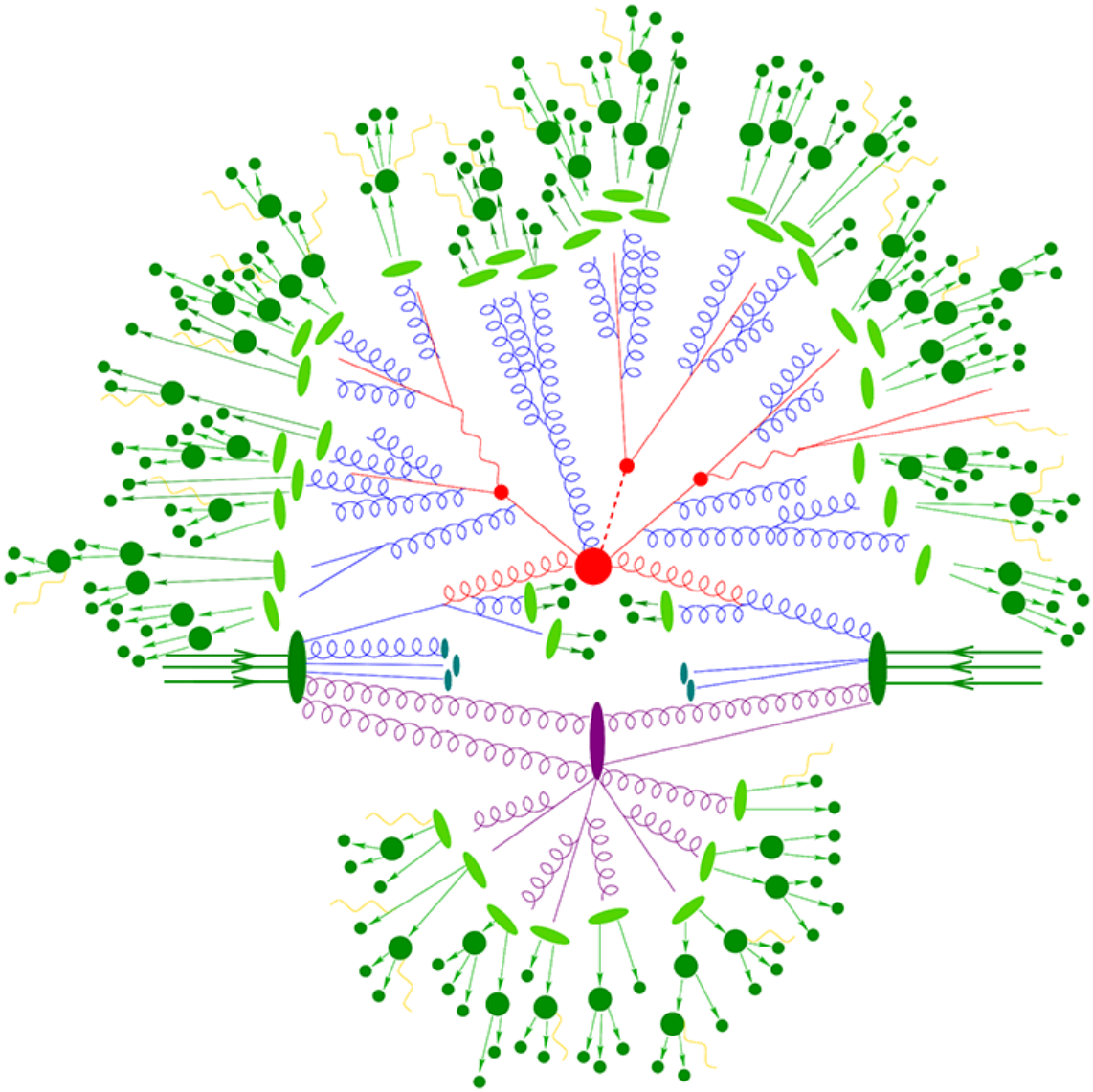


Figure 6.1: Cartoon depiction of the factorised stages in MC generation. Incoming hadrons are represented by the dark green lines with arrows pointing towards the centre of the figure. The hard scattering process is depicted in red, with UE/MPI in purple. The parton shower is shown in blue, the initial hadronisation is represented by light green, and the subsequent hadron decays and final state particles are shown in dark green, with electromagnetic radiation also represented in yellow. This particular example can be interpreted as containing the hard process $pp \to t\bar{t}H \to q\bar{q}bl\nu bb\bar{b}$, along with additional radiation and UE/MPI.

cross-sections can be split into perturbative and non-perturbative pieces. The non-perturbative effects are gathered into the parton distribution functions $f(x, \mu_f^2)$, which are summed across all partons i, j in Equation 6.1. μ_f is known as the *factorisation scale*, and is the scale chosen at which to separate the perturbative and non-perturbative terms. That is, interactions with energy below μ_f^2 are gathered in the parton distribution functions, and those above μ_f^2 are considered as part of the short-distance cross-section $\hat{\sigma}$, which is dependent upon the strong coupling α_S at the factorisation scale. Perturbative calculations can then be used to make precise predictions for a given process, such as the production and subsequent decay of top quark pairs in proton-proton collisions $\sigma_{t\bar{t}}$, and the sum of these processes (represented in Equation 6.1 as the sum of the indices i and j) is the total hadron-hadron cross-section.

6.2 Parton Distribution Functions

Inside a hadron there exists not only the *valence quarks* – *uud* in the case of the proton – which are typically thought to make up the particle, but also a *parton sea* of quarks and gluons which carry some fraction of the total hadron energy. The distribution of the energy inside the hadron is described by the *parton distribution function* (PDF). The DGLAP equations [166] describe the energy dependence of the PDFs, but the fraction x of the total hadron momentum P contained within each individual parton p is not analytically calculable. Instead, the distribution of momentum between the valence quarks, sea quarks, and sea gluons, can be parameterised and fit to data from experiment. Many such experiments exist, notably the HERA collider [167] located at DESY, Hamburg, where the ZEUS [168] and H1 [169] experiments made detailed studies of proton structure in electron-proton collisions. Indeed, experimental data from the LHC can also be included in these fits, where for example results such as those presented in Chapter 7 are particularly sensitive to the high-x gluon PDF [170]. The parton distribution functions are thus key ingredients in the MC simulations of pp collisions at the LHC.

6.3 Matrix Element Generation

In quantum field theory, the S-Matrix is the unitary operator which encodes the information on the scattering processes, and the elements of this matrix $\mathcal{M}$ are the scattering amplitudes – the sum of the Feynman diagrams for a given process N. The total cross-section for this process, with n particles in the final state plus up to k additional emissions and l loops per diagram, is given by [166]

$$\hat{\sigma} = \sum_{k=0}^{\infty} \int_{n+k} d\Phi_{n+k} \left| \sum_{l=0}^{\infty} \mathcal{M}_{n+k}(\Phi_{n+k}) \right|^2 . \tag{6.2}$$

Here, the phase space element $d\Phi$, ignoring normalisation factors, is given by

$$d\Phi_{n+k} \sim \prod_{f=0}^{n+k} \frac{d^3\boldsymbol{p}_f}{2E_f}, \tag{6.3}$$

with E_f and $\boldsymbol{p}_f$ the energy and 3-momentum of the particle f respectively.

The full calculation of Equation 6.2 is not possible analytically, and so instead is performed using perturbation theory. A leading order calculation is one in which k and l are restricted only to $k = l = 0$, such that the higher order Feynman diagrams with additional emissions and loops are not included. Including terms with $0 < k < N$ means including the Feynman diagrams with up to N additional real emissions, while full NLO calculations will include the $k = 1$ diagrams as well as interference with loops. These calculations are typically performed at NLO in packages such as POWHEG [171] or aMC@NLO [172], while the SHERPA [173] package is capable of calculations at NLO, with one emission at NLO + up to 4 additional emissions at LO.

6.4 Parton Shower

After the hard process has been generated at fixed order, the subsequent evolution of the event is modelled using a *parton shower*, which approximates corrections from higher order emissions in order to model the complete final state. Each parton in the initial and final state has some non-zero probability at the scale Q to undergo bremstrahlung radiation and branch into two particles. The parton shower calculates these probabilities and approximates the underlying physics process down to some cut off scale, typically around 1 GeV, meaning infrared and collinear divergences are avoided by including only resolvable branchings. The evolution of these branches is given by the *Sudakov Form Factor*

$$\Delta(t_0, t) \equiv \exp\left[-\int_{t_0}^{t} dt' \int dz \frac{\alpha_S}{2\pi} P(z)\right]. \tag{6.4}$$

Here the integrations are performed over the ordering variable t, and the relative momentum fraction z between the resultant partons. The order variable is typically a function of p_T or angle, and controls the sequence in which the splittings are calculated in the shower (and is thus analogous to a "time", hence the label t). $P(z)$ is the Altarelli–Parisi splitting function [174], which describes the quark and gluon splitting probabilities. These integrations are usually performed numerically using the Sudakov Veto Algorithm [175]. This algorithm approximates the integration of an arbitrary function $f(t)$ using a new function $g(t)$, which is easily calculable and for which $g(t) > f(t)\ \forall\ t$. Then, by vetoing trial splittings at t_i with a probability $f(t_i)/g(t_i)$, the correct probabilities are recovered.

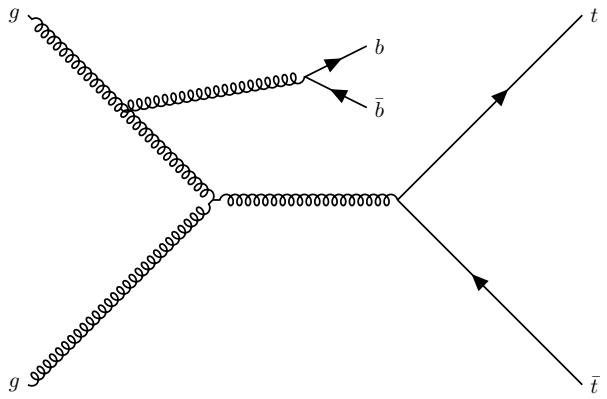


Figure 6.2: Example Feynman diagram for $t\bar{t} + b\bar{b}$ production.

6.4.1 Enhancing $g \to b\bar{b}$ Splitting in PYTHIA8

In many cases in ATLAS, physics measurements/searches are performed in very extreme corners of phase-space, where events are predicted to be very rare. One such example of this is the process $t\bar{t}H, H \to b\bar{b}$, where not only is the signal cross-section extremely small compared to the background, but this background ($t\bar{t} + b\bar{b}$) is itself only an extreme corner of the total inclusive $t\bar{t}$ cross-section. An example Feynman diagram for $t\bar{t} + b\bar{b}$ production is shown in Figure 6.2. In most cases, the $t\bar{t}$ process is generated inclusively, and additional b-jets arise primarily from the parton shower via rare $g \to b\bar{b}$ splitting. In order to generate sufficient statistics for this process, ATLAS uses a filtering strategy – generating inclusively, and saving and simulating only a subset of events which contain energetic B-hadrons in the final state. The filtering efficiency for such samples used in Chapters 8 and 9 is very low – as low as 0.1%, in the case of dilepton $t\bar{t} + b\bar{b}$ – meaning a huge amount of CPU time is wasted in generating events which fail this filter. While the detector simulation step takes up the most CPU time in general – averaging 245 s per event ATLAS wide, compared to 80 s per event for the event generation – particularly extreme phase spaces such as these can also significantly benefit from new techniques to increase the filter efficiency. Indeed, the MC statistics were one of the largest uncertainties on the published measurement in Chapter 8, and thus improved strategies are necessary particularly considering the increased luminosity foreseen in future data taking periods. One such way to more efficiently generate MC statistics in the relevant regions of phase-space is via Sudakov Veto Algorithm enhancements [176].

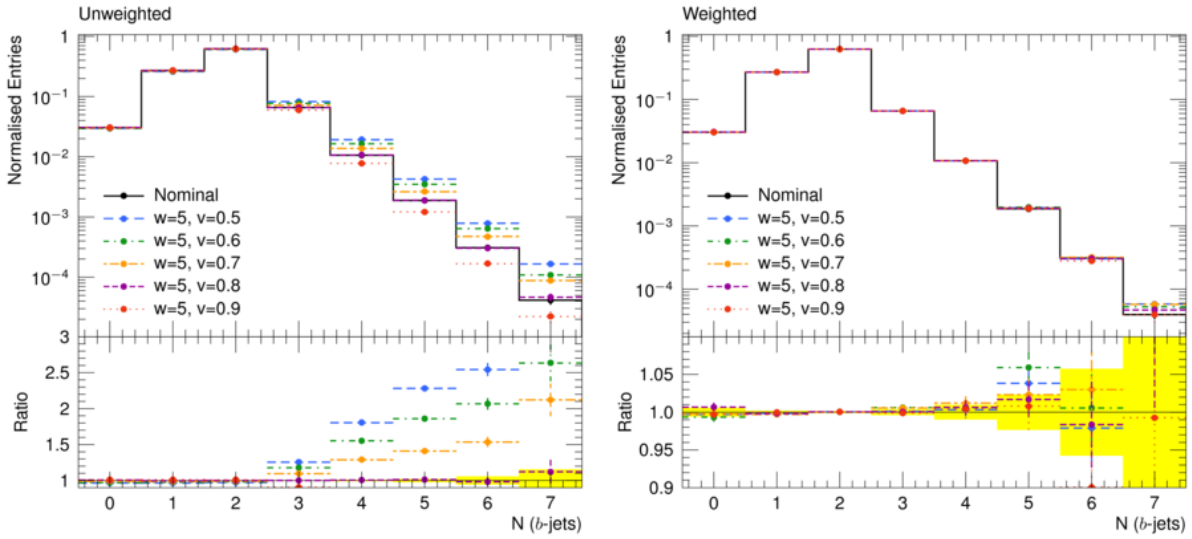
Since the Sudakov Veto Algorithm requires the selection of an arbitrary function $g(t)$, for which the only requirement is that $g(t) > f(t)\ \forall\ t$, then we can add an extra factor w onto this arbitrary function for particular branchings, such that $g_{\text{biased}} = w \times g(t)$. This has the result of making this branching w times more likely to occur, thus oversampling that region of phase space. The no-branching probability is maintained by another parameter v, which should be in the range $0 < v \leq 1$. By vetoing the splitting with probability v and giving it an event weight

> 1, or otherwise accepting the event and weighting it down by a factor < 1, kinematic closure can be maintained while still generating additional physics events in the interesting region of phase space.

Using this technique, and patching Pythia8 [158] to have direct access to the QCD branching $g \to b\bar{b}$, it is possible to enhance the production rate of $t\bar{t} + b\bar{b}$ events while maintaining kinematic closure. To show this, 1 million $t\bar{t}$ MC events are generated at NLO with Powheg, and subsequently showered using a patched version of Pythia8 v8.230. The same events are showered using the nominal settings (corresponding to $w = 1$, $v = 0.0$) as well as for a fixed value of $w = 5$ and a range of v parameters[1]. The case of $w = 1 - 1/v$ reduces back to the nominal setting of no enhancement, and so for $w = 5$ and $v = 0.8$, the nominal distribution is recovered with event weights of unity. Events are analysed with the Rivet [177] toolkit. Figure 6.3a shows the effects of such an enhancement in the b-jet multiplicity distribution without weights, showing the enhanced rate of additional b-jet production as expected. For $w = 5$, values of $v > 0.8$ represent suppression of the splitting, as observed for the dotted red line representing $v = 0.9$, which produces fewer $t\bar{t} + b\bar{b}$ events. The case of $w = 5$ and $v = 0.8$ is shown in dashed purple, and these settings reproduce the nominal distribution without event weights applied as expected. If $v < 0.8$, the $g \to b\bar{b}$ splitting is enhanced, with $v = 0.5$ producing approximately twice as many events with 4 b-jets, suggesting the filter efficiency of $t\bar{t} + b\bar{b}$ production could be increased by approximately a factor of two via this method. Figure 6.3b shows the same distribution with the event weights applied, where it is observed that kinematic closure is maintained within the statistical uncertainties of the samples. Further distributions related to the b-jet and $t\bar{t}$ kinematics are shown in Figures 6.4 and 6.5, which show that this kinematic closure holds in other distributions. An alternative way to evaluate the effectiveness of the technique is on the relative size of the statistical uncertainty for an inclusive $t\bar{t}$ sample on a $t\bar{t} + b\bar{b}$ selection. This is shown in Table 6.1, which utilises the Rivet routine developed for the ATLAS $t\bar{t} + b\bar{b}$ analysis performed at $\sqrt{s} = 8$ TeV [178], modified for the samples at $\sqrt{s} = 13$ TeV and to incorporate a single lepton selection. The statistical error on the nominal sample is 28.7% and 7.7% in the dilepton and lepton+jets selections respectively. By turning on the enhancement, this can be reduced to as little as 20.5% and 6.0%, a relative improvement of 40% and 28% respectively.

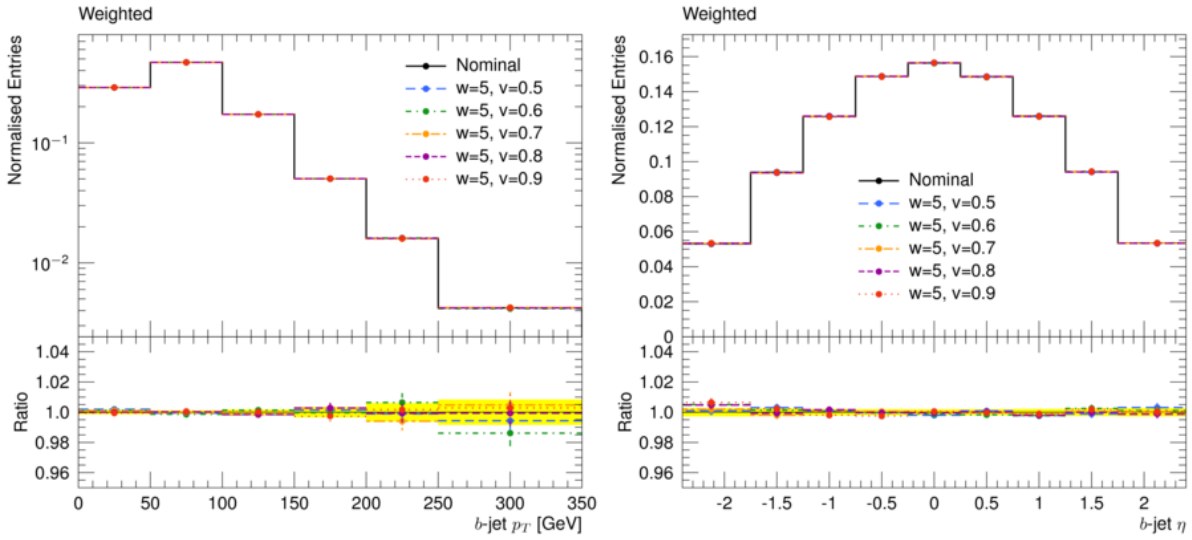
Despite the good kinematic closure observed, it is important to understand the weights that result from these enhancements. It is undesirable to have events with very large weights, as these can cause large fluctuations after fiducial selections are applied, and the overall statistical power of a sample reduces if the spread of weights increases – so there is a trade off in statistics to be considered. Table 6.1 demonstrates this: the most extreme enhancement of $w = 5, v = 0.5$ does not improve the statistical error relative to the settings $w = 5, v = 0.6$, reflecting the loss in

[1]The analogous study, varying w and fixing v, produces similar results, as it is only the relative values of w and v that are important.



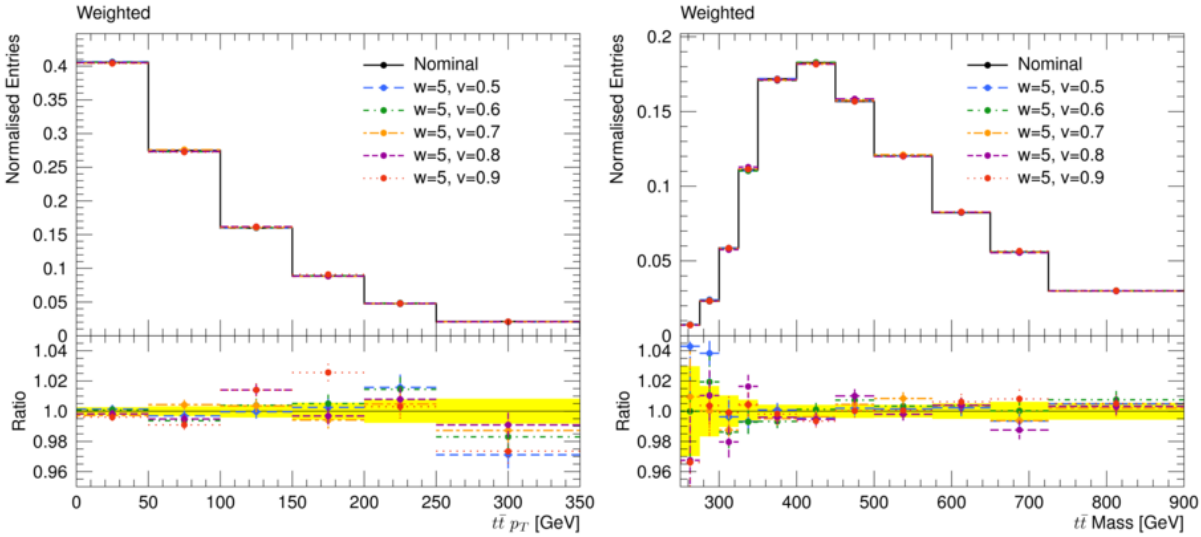
(a) (b)

Figure 6.3: Distributions of the number of b-jets in the generated $t\bar{t}$ samples, (a) without event weights applied, demonstrating the effect of the enhancement; and (b) with event weights applied, demonstrating good kinematic closure.



(a) (b)

Figure 6.4: Kinematic distributions of the b-jets in the generated $t\bar{t}$ samples with event weights applied, demonstrating good kinematic closure.



(a) (b)

Figure 6.5: Kinematic distributions of the $t\bar{t}$ system in the generated $t\bar{t}$ samples with event weights applied, demonstrating good kinematic closure.

Table 6.1: The cross-section and corresponding statistical errors for the $t\bar{t}$ + $b\bar{b}$ process, in both dilepton and lepton+jets selections, of 1 million inclusive $t\bar{t}$ events generated with various enhancement settings. The observed cross-sections are in good agreement, with the statistical error reducing as enhancement increases, as expected.

		Dilepton Selection		Lepton+Jets Selection	
w	v	σ [fb]	Error [%]	σ [fb]	Error [%]
1	0.0	41.6	28.7	505.6	7.7
5	0.9	45.4	38.4	468.9	10.9
5	0.8	44.6	26.4	531.7	7.6
5	0.7	38.1	25.2	509.2	6.5
5	0.6	45.0	20.5	517.8	6.0
5	0.5	37.1	20.5	501.1	6.0

statistical accuracy due to the spread of MC weights. As the enhancement becomes larger, the spread of MC weights becomes wider, and thus also the number of events with a large weight increases. The exact choice of w and v used for any samples produced with this technique must thus be carefully chosen depending on the intended application, to balance the size of the enhancement against the resulting weights distribution.

Though this study was performed for $t\bar{t} + b\bar{b}$ production, an identical technique can be used for $t\bar{t} + c\bar{c}$ production, which is also a highly relevant background to the $t\bar{t}H, H \to b\bar{b}$ analysis, by enhancing the $g \to c\bar{c}$ splitting. Indeed, any splitting can be enhanced in this manner, and this functionality has been implemented into PYTHIA8 v8.235 and onward following the results of this study.

6.5 Hadronisation

After the showering is complete, the resulting partons must undergo *hadronisation* into the bound baryonic and mesonic states that are ultimately observed in the detector, due to confinement. There are two models in common use today to describe this process: Lund string fragmentation [179] and the cluster model [180]. In the Lund string model, used in the PYTHIA event generators, the quarks are considered to be joined by a one-dimensional string about 1 fm in length, with a tension of around 1 GeV/fm. The gluons are considered as waves carrying energy along these strings. Because of the principle of quark confinement, the quarks moving apart causes an increase in the potential energy between them. This, along with the energy from the gluon waves, causes it to eventually become energetically favourable for the string to break and produce a new $q\bar{q}$ pair. These become the ends of two smaller strings, and in this way the energy of the collision is dissipated down until the final, detectable hadronic states. In contrast, the cluster model considers the gluons left over at the end of the parton shower to decay isotropically into $q\bar{q}$ pairs. These along with the other final quarks then cluster to create colour singlets, which can effectively be treated as an excited hadron resonance, which will in turn potentially decay further until the hadrons become stable. This model is used in the HERWIG event generator / parton shower package.

6.6 Underlying Event / Multi-parton Interactions

After a hard interaction between two incoming partons, the remaining quarks and gluons from the colliding hadrons are also then subject to secondary interactions which create additional activity known as the *Underlying Event* (UE). These remnants are likely no longer in color confined states and thus undergo further hadronisation. Additional interactions between remnant partons is referred to as *multi-parton interaction* (MPI), and these effects must also be accounted

for in the MC simulation. Measurements can be made of these processes using minimum bias data [181] and subsequently used to tune relevant MC parameters that describe this activity.

6.7 Detector Simulation

In order to compare directly the data and the MC, a detailed simulation of the effect of the interactions between the particles and the detector hardware must be included. These effects, accounting for the finite resolution of the various technologies utilised in recording the particle collisions, are applied to the final state particles output from the MC generators. The simulation of the full ATLAS detector is a highly complex process, and is often the dominant contribution to the total generation time of the samples. The interaction with the calorimeter is particularly computationally complex, and for this reason, two methods of simulation exist within ATLAS. The first, using GEANT4 [182], includes a full simulation of the interactions with every piece of the detector. This method is referred to as *full simulation (full-sim)*, and is the standard simulation method used in most MC samples. An alternative, faster simulation, referred to as *AtlFastII (AFII)* [183], uses a parameterised calorimeter description that drastically reduces the CPU overhead – by around an order of magnitude or more – in exchange for a poorer description of the calorimeter response. This effect can be mitigated via dedicated calibration to AFII MC, and this results in good agreement between data and MC at the level of the high level objects, though jet substructure remains poorly modelled.

6.8 Generated Samples

In the analyses presented in Chapters 7, 8, and 9, most processes are modelled with MC samples generated using the techniques described here. Since there are many similarities between the samples used in each analysis, they are described here in full, with any differences highlighted. In all cases, the top quark mass is set to $m_{\text{top}} = 172.5$ GeV.

$t\bar{t}$

The analysis in Chapter 7 uses a $t\bar{t}$ sample, generated at NLO using POWHEG-BOXV2 [184] and showered with PYTHIA6 [175], to model the signal process. This sample uses the CT10 [185] PDF set in the matrix element generator and the CTEQ6L1 [186] PDF set with the Perugia2012 [187] tune in the parton shower. The POWHEG parameter h_{damp}, which controls the scale of the hardest emission, is set to m_{top}. Two alternative samples are generated in which the factorisation and renormalisation scales are varied by a factor of two up or down, and the corresponding "radLo" and "radHi" variations of the Perugia tune are used in the parton shower. In the upward variation,

h_{damp} is also simultaneously varied by setting $h_{\text{damp}} = 2m_{\text{top}}$, while in the downward variation h_{damp} is not varied [188].

The nominal $t\bar{t}$ background in Chapters 8 and 9 is modelled with a similar setup, but using Pythia8v8.210 and the A14 tune [189] for the parton shower and hadronisation. In this sample the NNPDF3.0NLO PDF [190] is used in the matrix element generation with $h_{\text{damp}} = 1.5 \times m_{\text{top}}$. This value was chosen after comparison to 13 TeV data, including the results presented in Chapter 7, found it to give better data/MC agreement. This setup is also compared to the unfolded results of Chapter 7. The radiation variations also similarly use variations of a factor two in μ_f, μ_r and h_{damp}, along with the "Var3c" up/down variations of the A14 tune.

Further $t\bar{t}$ samples are produced using alternative generator setups, both for defining systematic uncertainties and for comparing to unfolded results in Chapter 7. Samples using Herwig++ [180] and Herwig7 [191], with the nominal Powheg settings plus $h_{\text{damp}} = m_{\text{top}}$ and $h_{\text{damp}} = 1.5 \times m_{\text{top}}$ respectively, offer a way to estimate the uncertainty due to choice of parton shower and hadronisation package. The Herwig++ sample, used only in Chapter 7, uses the CTEQ6L1 PDF and UE-EE-5 tune [189], while the Herwig7 sample used in all three analyses uses the H7-UE-MMHT tune [191]. Further samples are generated using MadGraph5_aMC@NLO [172] for the matrix element calculation, using the CT10 PDF set, and subsequently interfaced to the Pythia8, Herwig++, and Herwig7 setups already described. Finally, a sample is generated at NLO using Sherpa v2.2.1 [173], interfaced with OpenLoops [192], and using the NNPDF3.0NNLO PDF. Up to one additional emission is included in the calculation at NLO and up to four are included at LO.

All of these samples are normalised to the NNLO+NNLL cross-section $\sigma_{t\bar{t}} = 823^{+46}_{-51}$ pb, as calculated in the Top++2.0 program [193], with top quarks decayed using MadSpin [194] to preserve spin correlation effects.

$t\bar{t}$+Heavy Flavor

In the $t\bar{t}H$ analyses in Chapters 8 and 9, the $t\bar{t}$ backgrounds are further subdivided into components based on the flavor composition ($b-$ and $c-$jet) of additional radiation in the event. To increase the available MC statistics in these components, additional events are generated with the same setups as described above, with filters applied requiring such "heavy flavor" (HF) jets.

The samples described thus far utilise the 5-Flavor scheme [195], in which the quark masses (other than top) are set to 0, and the additional heavy flavor contributions come purely from the parton shower. Two additional samples are generated to assess the modelling of the $t\bar{t}$+HF processes. A $t\bar{t} + b\bar{b}$ sample is generated using Sherpa v2.2.1 + OpenLoops, with the CT10 PDF set in the 4-Flavor scheme. The use of the 4-Flavor scheme, in which the b-quark mass effects are included, ensures that the calculation is fully at NLO down to $p_T = 0$. Similarly, for $t\bar{t} + c\bar{c}$, a dedicated sample is generated using MadGraph5_aMC@NLO+Herwig++ and the CT10F3 [196] PDF that uses the 3-Flavor scheme [197] in which c-quark mass effects are also

included.

$t\bar{t}H$

The signal process in the $t\bar{t}H$ analysis presented in Chapter 8 is modelled using an NLO matrix element generated with MadGraph5_aMC@NLO and subsequently showered using Pythia8 with the A14 tune and NNPDF3.0NLO PDF. In Chapter 9, a sample generated with Powheg + Pythia8, with the same settings as for the $t\bar{t}$ samples, is used as nominal with the MadGraph5_aMC@NLO sample used as an alternative sample to define a systematic uncertainty on the ME generation. The analyses in both chapters also include a systematic uncertainty on the parton shower, with Chapter 8 and Section 9.1 using a MadGraph5_aMC@NLO+Herwig++ sample and Section 9.3 using MadGraph5_aMC@NLO+Herwig7. All of these samples are normalised to the NLO QCD + NLO EW cross-section of 507fb [6, 198–202], decayed with MadSpin, and utilise Higgs boson branching fractions calculated using HDecay [203].

The contribution from the $t\bar{t}H$ process is negligible in the analysis presented in Chapter 7 and so these samples are not included there.

$t\bar{t}$+X

Associated production with other processes, such as $t\bar{t}$ + W, $t\bar{t}$ + Z, and $t\bar{t}$ + $t\bar{t}$ production, are considered as backgrounds in all the analyses presented. In Chapter 7, $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}WW$processes are generated with MadGraph5_aMC@NLO+Pythia8 v8.186, using A14 and NNPDF2.3NLO as the PDF. Chapters 8 and 9 use an updated version of this setup, using Pythia8 v8.210 and NNPDF3.0NLO. All of these samples are normalised to their NLO cross-sections [172]. In Chapter 8 and Section 9.1, production of $t\bar{t}t\bar{t}$ is considered as a background, and is generated using the same setup. The effect of these samples is small, and is neglected in the studies presented in Section 9.3.

Single Top

Processes in which a single top quark is produced are generated using Powheg-Box v1 interfaced to Pythia6 and normalised to NNLO cross-sections [204–206]. Wt and s-channel production, generated in the 5-flavor scheme, use the CT10 PDF, while t-channel production is generated in the 4-flavor scheme and so uses CT104F. Top quarks are decayed with MadSpin, and overlap between $t\bar{t}$ and Wt production is handled with the diagram removal scheme [207] in the nominal sample, with a comparison to the alternative diagram subtraction scheme considered as a systematic uncertainty in all the analyses presented. An alternative sample, using Powheg + Herwig++, is used to define a systematic uncertainty on the PS model used for Wt and t-channel in the analysis presented in Chapter 8.

W/Z+Jets and Diboson

Samples of vector boson production are generated using SHERPA, with matrix elements from COMIX [208] and OPENLOOPS interfaced to the SHERPA parton showering algorithm with the CT10 PDF set. W+jets and Z+jets processes are normalised to the NNLO cross-sections [209], while diboson production is already generated at NLO and no additional normalisation is applied. In Chapter 7 SHERPA v2.1.1 is used, which is updated to v2.2.1 for Chapters 8 and 9.

Dijet

Dijet MC samples are generated at leading order using PYTHIA8 v8.230 with the NNPDF2.3 PDF, HERWIG7 with the NNPDF3.0 PDF, and SHERPA2.2.1 with the CT10 PDF. These samples are primarily used to evaluate the systematic uncertainties on the reclustered jet substructure variables, as described in Section 9.2.

Chapter 7

Differential Cross-Section Measurements of Boosted Top Quarks

The chances of finding out what's really going on in the universe are so remote, the only thing to do is hang the sense of it and keep yourself occupied.

– Douglas Adams

When Run 2 of the LHC began in 2015, at an increased centre-of-mass energy of $\sqrt{s} = 13$ TeV, it was important to carry out early measurements of SM processes, in order to understand physics at this new energy frontier and to test the validity of the Standard Model. This chapter presents a differential measurement of boosted top quark kinematics using 3.2 fb^{-1} of $\sqrt{s} = 13$ TeV data recorded in 2015, published in [1]. Two distributions are measured: the hadronic top quark transverse momentum $p_T^{\text{t,had}}$and absolute rapidity $\left|y^{\text{t,had}}\right|$. The top quark p_T distribution was observed to be poorly modelled by MC in Run 1 by both ATLAS [210,211] and CMS [212,213], motivating a new analysis at $\sqrt{s} = 13$ TeV. Results can be in terms of either *absolute* or *relative* cross-section, depending on whether or not the distributions are normalised to unity (relative) or not (absolute). This normalisation removes a degree of freedom and reduces the effect of many systematic uncertainties, but also removes information on the total magnitude of the cross-section. Thus, results are presented using both strategies.

7.1 Event Selection

The analysis targets highly boosted lepton+jets events, in which one top quark decays leptonically, while the second top quark decays hadronically. High p_T events are targeted by reconstructing the hadronically decaying top quark as a single large radius jet, utilising physics objects as described in Chapter 5.

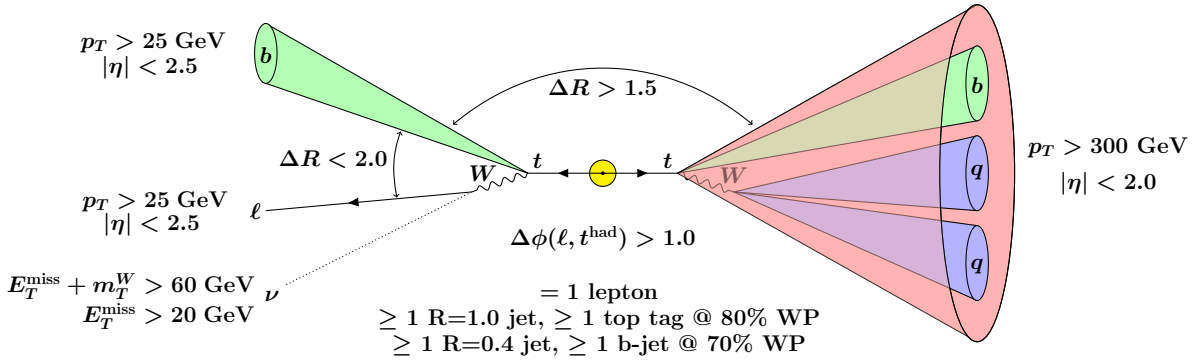


Figure 7.1: Diagram summarising the topology and event selection of the analysis.

7.1.1 Detector Level Selection

The event pre-selection requires exactly one electron or muon which pass the single lepton triggers described in Section 4.2.5, at least one top-tagged (at the 80% working point) large jet, and at least one small jet well separated from the large jet ($\Delta R > 1.5$, referred to as *additional* small jets), all as defined in Section 5. Further requirements on $E_T^{\text{miss}} > 20$ GeV and $E_T^{\text{miss}}+m_T^W > 60$ GeVare used to suppress the multijet background. If there is more than one top-tagged large jet, the leading jet in p_T is taken as the hadronic top candidate. The top jet and the lepton must be separated by $\Delta\phi > 1.0$, in order to prevent double counting of calorimeter clusters. Subsequent requirements are made on the topology of the event to maintain acceptance of $t\bar{t}$ events while rejecting backgrounds: at least one of the additional small jets must be within $\Delta R < 2.0$ of the lepton, and either one of these additional jets or one of the small jets reconstructed within the top jet ($\Delta R < 1.0$) must be b-tagged at the 70% working point. All of these requirements have been optimised using the MC predictions to balance signal purity against selection efficiency, and are summarised in Figure 7.1. These selections, applied to both data and MC, are referred to as the *detector level* selection. The event yields at detector level are shown in Table 7.1, with the data and MC compatible within uncertainties.

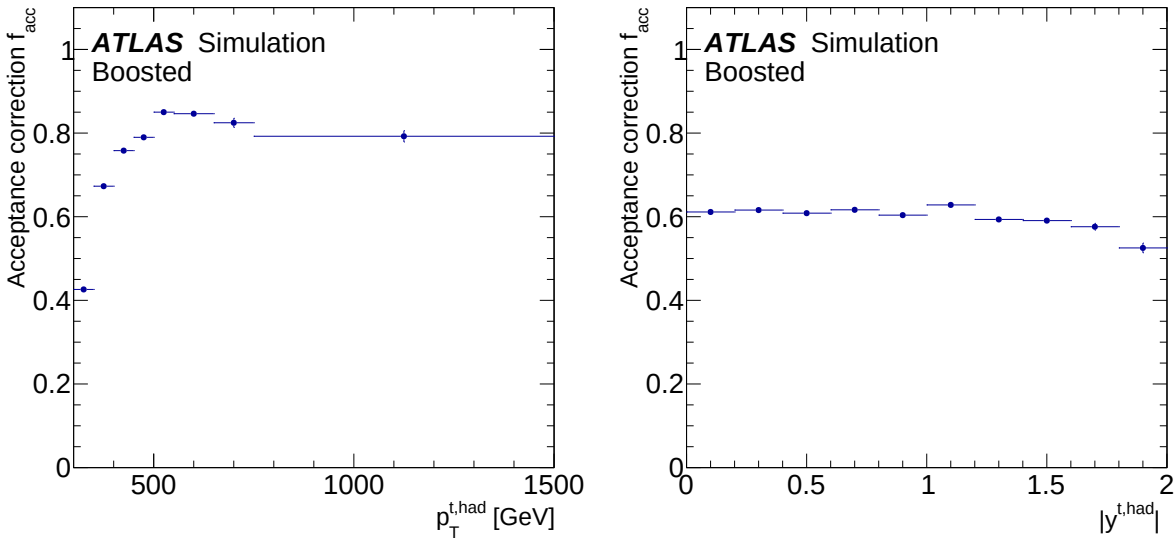
7.1.2 Particle Level Selection

In order to allow comparison to state-of-the-art MC predictions, the data is unfolded (see Section 7.3) to *particle level*, the stage of the MC simulation before detector effects are included. Thus, an equivalent selection to the detector level selection is required. Leptons and small jets are required to have $p_T > 25$ GeV and $|\eta| < 2.5$, large jets are required to have $p_T > 300$ GeV and $|\eta| < 2.0$, and the E_T^{miss} and m_T^W requirements are identical to detector level. Small jets are considered b-tagged if there is a ghost matched [146] B-hadron in the MC event record. A large jet is considered top tagged if it has $m > 100$ GeV and $\tau_{32} < 0.75$, with these values chosen to minimise bin-to-bin migrations while also simplifying from the p_T dependent cuts made at

Table 7.1: Event yields after the detector level selection. The uncertainties include both statistical and systematic uncertainties. The data yield is lower than the central value of the total prediction, but compatible within the uncertainties.

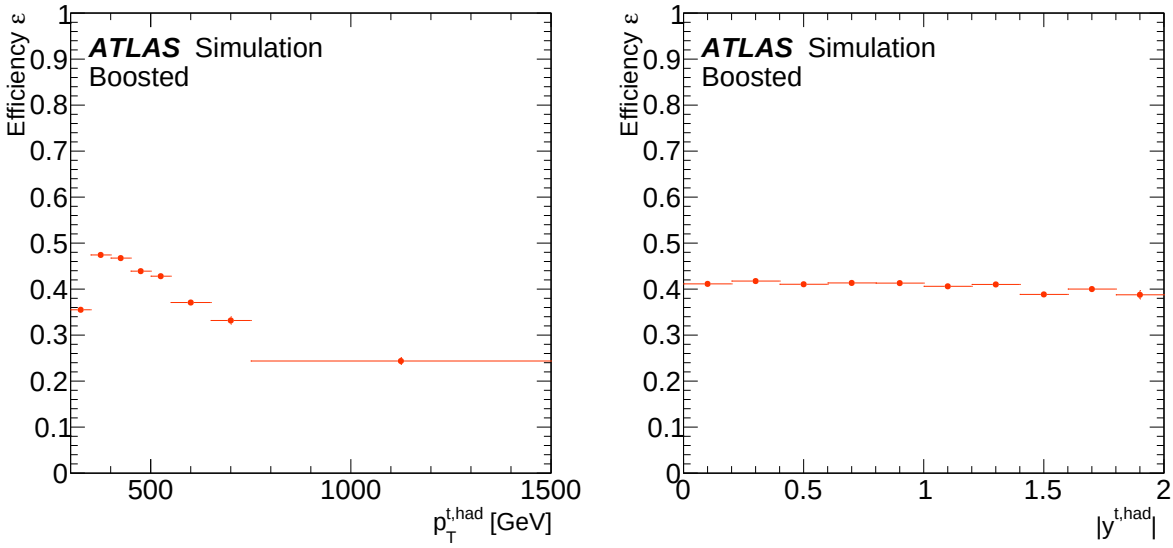
Process	Events
$t\bar{t}$	7000 ± 1100
W+jets	500 ± 200
Single top	500 ± 80
Multijets	300 ± 80
$t\bar{t}V$	70 ± 10
Z+jets	60 ± 40
Diboson	60 ± 10
Total Prediction	8300 ± 1300
Data	7368

detector level. This leaves a simply defined fiducial phase space independent of the experimental apparatus. The resulting acceptance and efficiency corrections, as defined in Section 7.3, are shown in Figures 7.2 and 7.3, and the migration matrices between the detector and particle level selections are shown in Figure 7.4. The acceptance and efficiency corrections are flat in $\left|y^{\text{t,had}}\right|$, and show the expected shape in $p_T^{\text{t,had}}$, reflecting the difficulty in tagging only moderately boosted top quarks and the loss in efficiency at very high p_T for things such as b-tagging. The migration matrices show good diagonality, with > 50% in each diagonal bin in p_T and > 85% in rapidity.



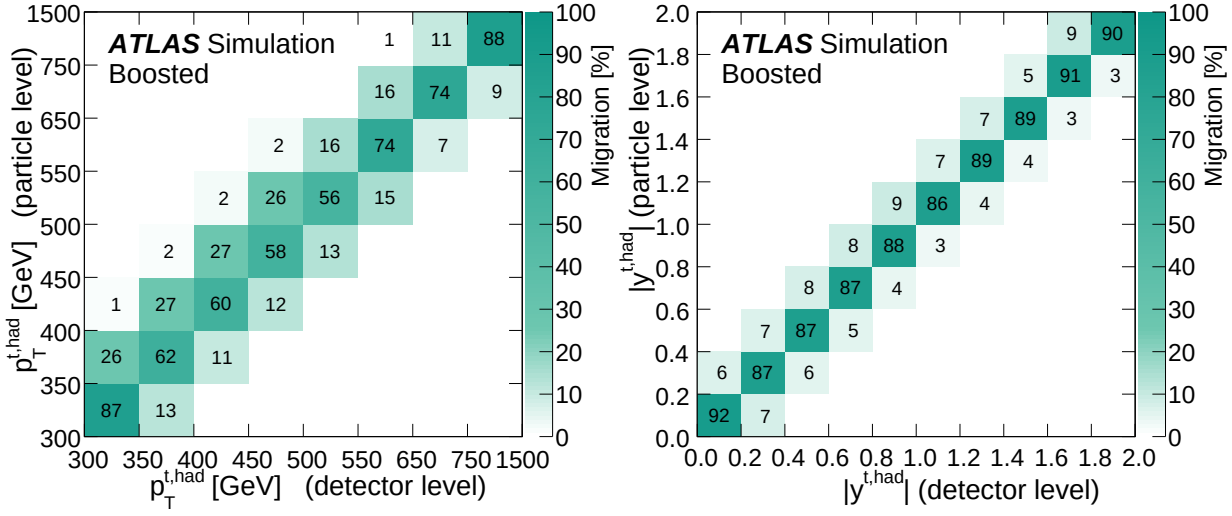
(a) (b)

Figure 7.2: The unfolding acceptance correction factors, as a function of (a)) $p_T^{\text{t,had}}$ and (b) $\left|y^{\text{t,had}}\right|$.



(a) (b)

Figure 7.3: The unfolding efficiency correction factors, as a function of (a)) $p_T^{t,had}$ and (b) $|y^{t,had}|$.



(a) (b)

Figure 7.4: The unfolding migration matrices, normalised by row, for (a) $p_T^{t,had}$ and (b) $|y^{t,had}|$.

7.2 Background Estimation

Background processes are modelled by a mixture of MC and data driven methods. The single top, Z+jets, diboson, and $t\bar{t}V$ backgrounds use purely MC based predictions, as described in Section 6.8. Two backgrounds are estimated using data driven methods: the W+Jets MC prediction is corrected for known mismodelling using data driven scale factors, while the QCD Multijet background is estimated with a purely data driven methodology due to the difficulty in efficiently modelling this process in MC.

7.2.1 Data Driven W+Jets Scale Factors

The W+jets MC sample is assigned a scale factor, based on data, to improve the accuracy of the prediction for what is the largest source of background events. There are two components to this scale factor. First, the charge asymmetry of $W^{\pm}$ production in pp collisions – with twice as many valence up quarks than down quarks in the collision, and no valence anti-quarks, $u\bar{d} \rightarrow W^+$ production is favoured over $\bar{u}d \rightarrow W^-$ – is corrected for. The ratio

$$r = \frac{N_{W^+}}{N_{W^-}} = \frac{\sigma(pp \rightarrow W^+)}{\sigma(pp \rightarrow W^-)}, \tag{7.1}$$

where N_W^+ is the number of positively charged events and N_W^- is the number of negatively charged events, is well understood theoretically and in fact has smaller uncertainties than the W+jets cross-section [214]. This can thus be used to derive a scale factor S to scale the MC prediction

$$S = \frac{N_W^{\text{data}}}{N_W^{\text{MC}}}, \tag{7.2}$$

where N_W^{MC} is the number of W events in MC, and N_W^{data}, the number of W events in data, is estimated using the charge asymmetry via the equation

$$N_W^{\text{data}} = N_{W^+}^{\text{data}} + N_{W^-}^{\text{data}} = \frac{N_{W^+}^{\text{MC}} + N_{W^-}^{\text{MC}}}{N_{W^+}^{\text{MC}} - N_{W^-}^{\text{MC}}}(D^+ - D^-) = \frac{r_{\text{MC}} + 1}{r_{\text{MC}} - 1}(D^+ - D^-). \tag{7.3}$$

Here, r_{MC} is the ratio as calculated from MC, while $D^{+(-)}$ is the number of positive (negative) charge data events, after subtracting the charge-asymmetric contributions from other processes such as single top, calculated from MC. Charge symmetric contributions are then implicitly removed by taking this difference, leaving only the W+jets charge asymmetric contribution. Thus, assuming that the asymmetry is equal in MC and data, scaling this number by the MC estimate of the charge asymmetry gives the number of W+jets events in the data. The resulting scale factor can then be applied to the MC estimate for the total W+jets yield to improve the description of the data.

The relative fraction of W events produced in association with different jet flavors ($W + b\bar{b}$,

$W+c\bar{c}$, $W+c$, W+light) are known to be poorly modelled in the MC, and so dedicated scale factors for these fractions are also derived, following the procedure used in [215]. This is performed using events in an orthogonal control region to the nominal selection by requiring exactly two small jets and at least one b-tag. Three scale factors, $K_{b\bar{b},c\bar{c}}$, K_c, and K_l, are derived via a system of linear equations. The ratio of $W + b\bar{b}$ and $W + c\bar{c}$ is kept constant, and thus only a single scale factor is derived for both. The equations are constrained by the number of data events containing negatively or positively charged leptons and the requirement to keep the overall normalisation constant, and the scale factors are extrapolated to the nominal region assuming constant relative rates. Since this procedure depends on the charge asymmetry scale factor, the procedures for both are iterated until there is agreement between data and MC yields in the control region at the per mille level.

7.2.2 Data Driven QCD Multijet Estimate

The other background estimated via a data driven method is the QCD multijet background in which a non-prompt lepton or misidentified jet passes the lepton selection criteria. Because of the number of possible processes that result in this signature, MC estimation is very CPU intensive and impractical. Instead, this background can be estimated using data by estimating how often a "fake" lepton is reconstructed. For this, the "matrix method" [216] is used. This method uses differing selection criteria in the lepton reconstruction to select "tight" and "loose" leptons. An electron (muon) is considered tight if it passes the requirements described in Section 5.2.1 (5.2.2). A loose electron (muon) must pass the mediumLH (medium) ID requirements, as described in [136] ([134]), and has no isolation requirement applied in either case. The number of such events is then governed by the simultaneous equations

$$N^l = N^l_r + N^l_f, \tag{7.4}$$

$$N^t = \epsilon_r N^l_r + \epsilon_f N^l_f. \tag{7.5}$$

Here, $N^{l(t)}$ represents the number events with a loose (tight) lepton, with the subscripts r and f corresponding to real and fake leptons respectively. The efficiency $\epsilon_{r(f)}$ is the fraction real (fake) loose leptons which also pass the tight selection. The number of events containing tight fake leptons which pass selection can then be calculated as

$$N^t_f = \frac{\epsilon_f}{\epsilon_r - \epsilon_f}(\epsilon_r N^l - N^t), \tag{7.6}$$

where ϵ_r is derived from the data using the tag and probe method [124], while ϵ_f is measured using control regions containing a large number of fakes. These efficiencies are parameterised in terms of the event kinematics: electron p_T, η, N(b-tags), and $\Delta\phi$(lep, E_T^{miss}) in the electron

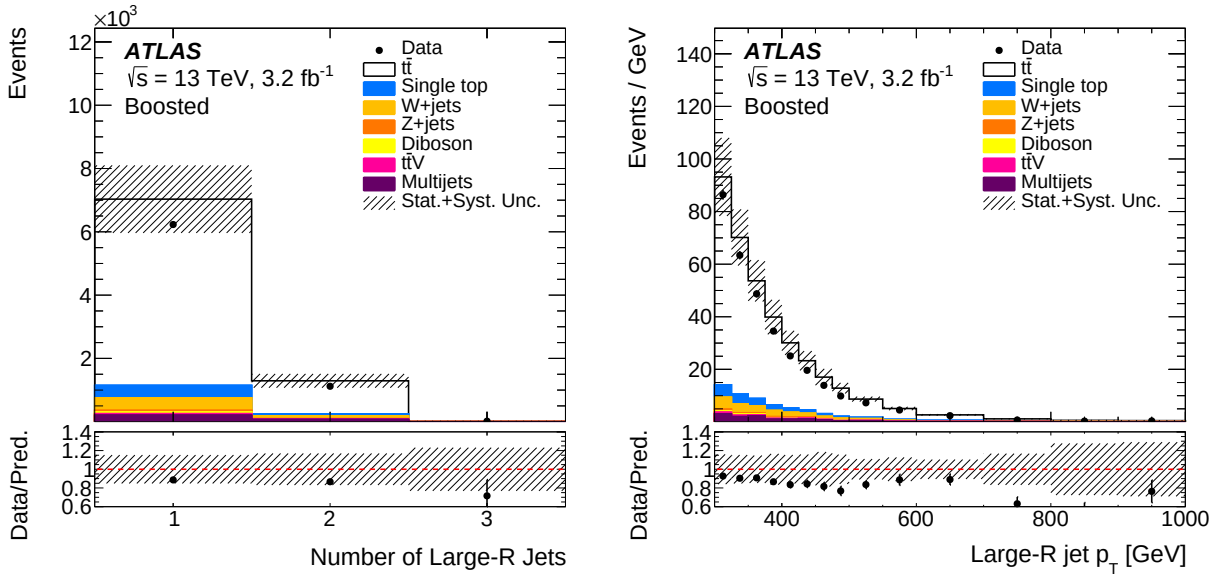
channel, and muon p_T, E_T^{miss}, and $\Delta\phi(\text{lep}, E_T^{\text{miss}})$ in the muon channel. Thus each data event is given a "fake weight" given by

$$w = \frac{\epsilon_f}{\epsilon_r - \epsilon_f}(\epsilon_r - \delta), \tag{7.7}$$

where $\delta = 1$ if the lepton passes the tight selection and $\delta = 0$ if not. One feature of this method is the possibility to obtain a negative weight, which can also lead to negative overall yields in some bins of some distributions. In any place where this occurs the fake weight is set to zero, and this can thus lead to a slightly different estimate in different variables. The yield in Table 7.1 uses the integral of a single binned distribution and is thus not necessarily the same as those in the distributions to be unfolded.

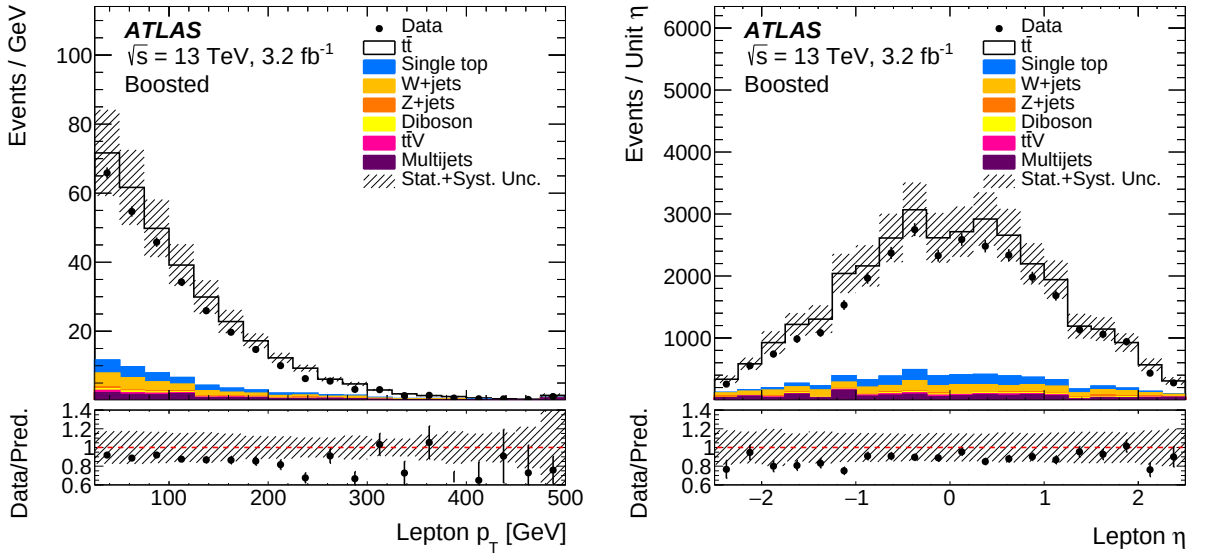
7.2.3 Control Plots

In order to check the validity of the signal and background predictions, many comparisons are made to the real data at detector level. As seen in Table 7.1, there is an overall overestimation of the data, though the prediction is compatible within the uncertainties. Figure 7.5 shows comparisons of the p_T distribution and number of large jets, showing reasonable agreement excepting this overall normalisation difference. Similarly, Figure 7.6 shows lepton kinematics, and Figure 7.7 shows E_T^{miss} and m_T^W. No significant mismodelling is seen in any of these distributions. Figure 7.8 shows the hadronic top candidate τ_{32} and mass distributions that are used in the top-tagging. τ_{32} peaks to the right of the distribution, consistent with the three-pronged topology expected from a top quark decay, and mass peaks are visible at both the top quark and W boson masses. This implies that in some cases, the hadronic top candidate selected likely is the hadronically decaying W boson, while the b-quark is not contained within the large jet. This is not a problem, as the migration matrices (shown in Figure 7.4) show good diagonality, demonstrating that this selection is still a good proxy to the top quark kinematics. Finally, Figure 7.9 shows $p_T^{\text{t,had}}$ and $\left|y^{\text{t,had}}\right|$, the distributions to be unfolded. Here, there are hints of a mismodelling in $p_T^{\text{t,had}}$, in a similar fashion to what was observed in the 8 TeV result [211], while $\left|y^{\text{t,had}}\right|$, which was not previously unfolded, is generally well modelled.



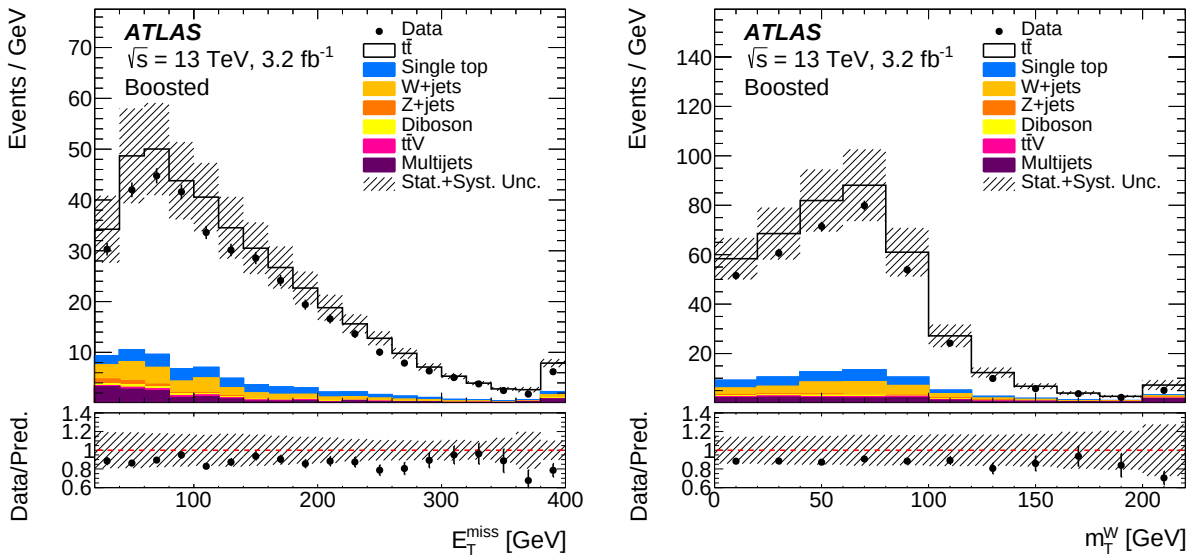
(a) (b)

Figure 7.5: Detector level plots of (a) the number of large jets, and (b) the large jet p_T. The top panel shows the data compared to the simulated signal and background expectation, while the ratio of Data/Prediction is shown in the bottom panel. The hatched areas show the statistical and systematic uncertainties on the measurement.



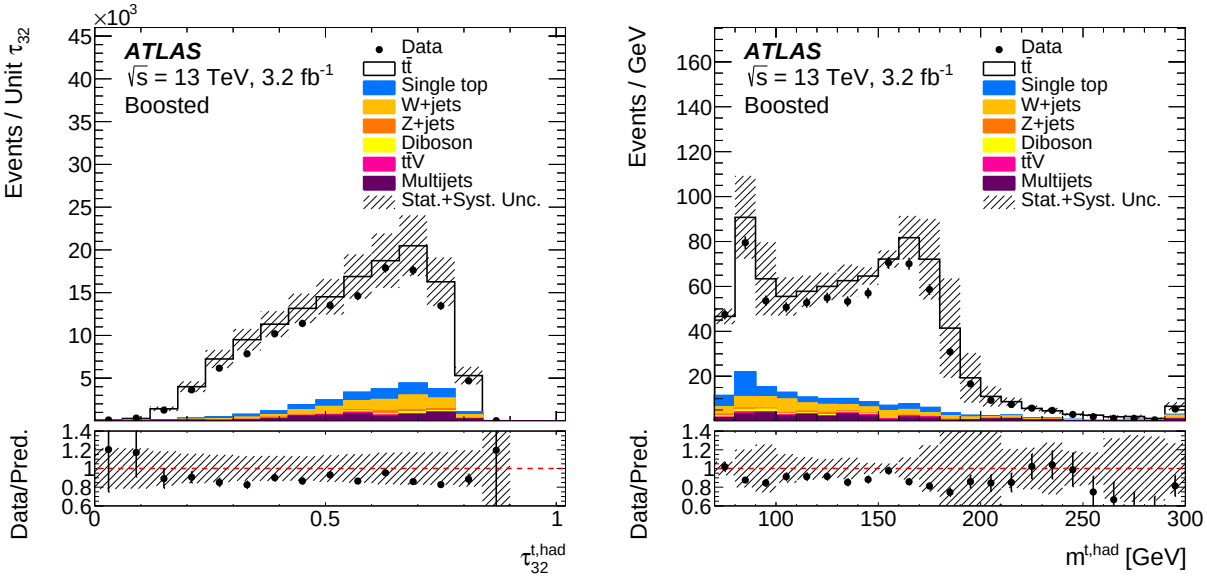
(a) (b)

Figure 7.6: Detector level plots of the lepton (a) p_T, and (b) η. The top panel shows the data compared to the simulated signal and background expectation, while the ratio of Data/Prediction is shown in the bottom panel. The hatched areas show the statistical and systematic uncertainties on the measurement.



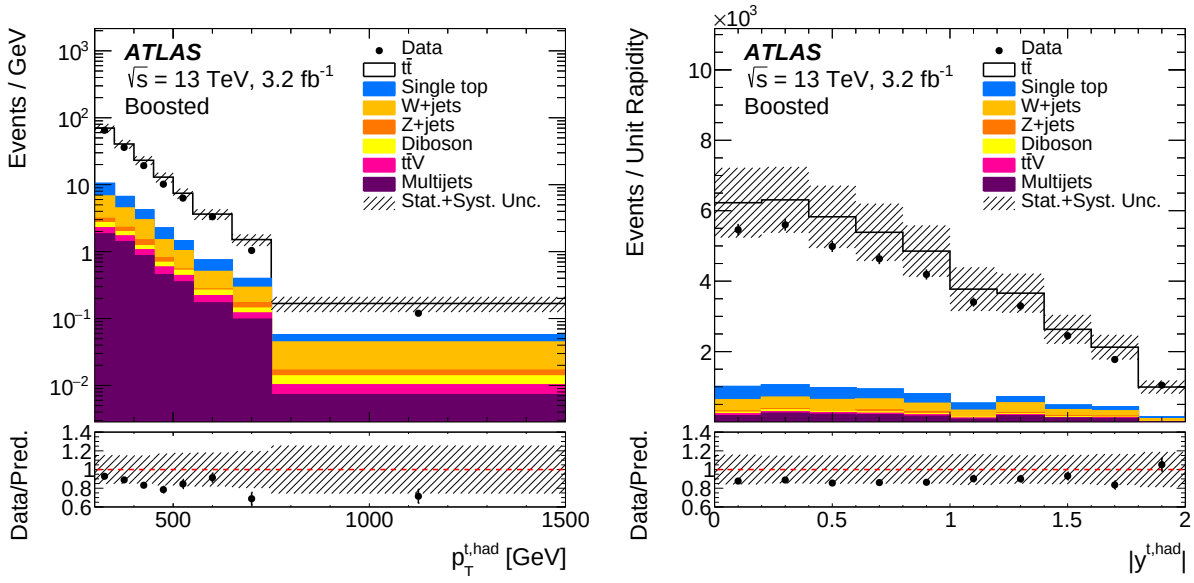
(a) (b)

Figure 7.7: Detector level plots of (a) E_T^{miss}, and (b) m_T^W. The top panel shows the data compared to the simulated signal and background expectation, while the ratio of Data/Prediction is shown in the bottom panel. The hatched areas show the statistical and systematic uncertainties on the measurement.



(a) (b)

Figure 7.8: Detector level plots of the hadronic top candidate jet (a) τ_{32}, and (b) mass. The top panel shows the data compared to the simulated signal and background expectation, while the ratio of Data/Prediction is shown in the bottom panel. The hatched areas show the statistical and systematic uncertainties on the measurement.



(a) (b)

Figure 7.9: Detector level plots of the hadronic top candidate jet kinematics, (a) $p_T^{\text{t,had}}$, and (a) $\left|y^{\text{t,had}}\right|$. The top panel shows the data compared to the simulated signal and background expectation, while the ratio of Data/Prediction is shown in the bottom panel. The hatched areas show the statistical and systematic uncertainties on the measurement.

7.3 Unfolding Tests

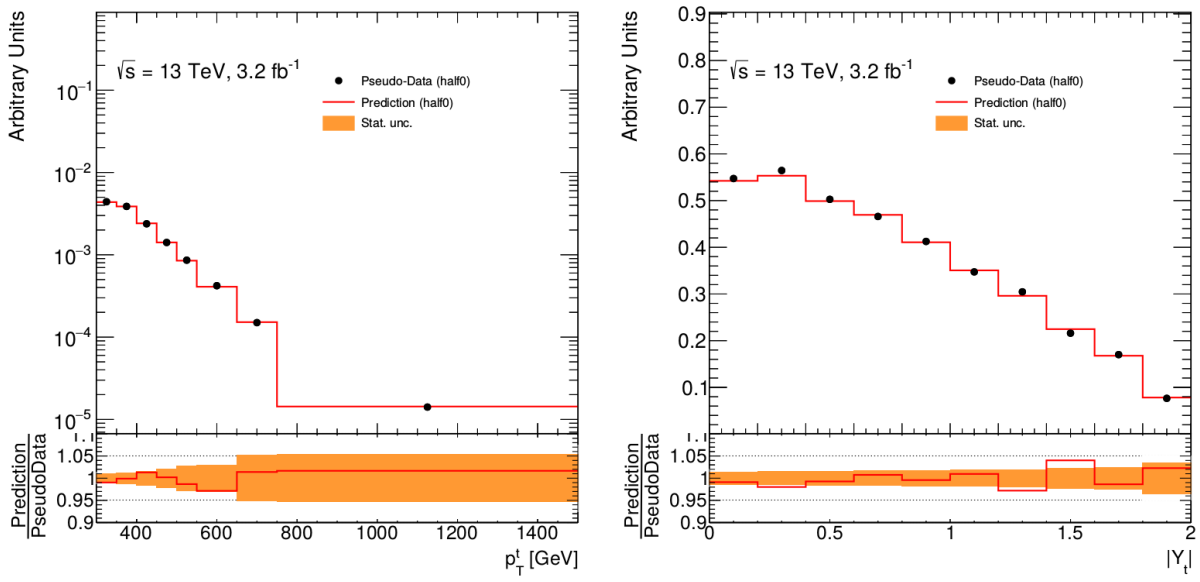
A number of tests of the unfolding procedure, as described in Section 3.2, are required in order to ensure no bias is introduced. Firstly, the number of iterations N used in the unfolding is tested, where the difference between $N = 3, 4, 5$ was found to be negligible in the final results. Thus, $N = 4$ is used throughout. A *closure* test checks that the unfolding correctly reproduces the particle level distribution in the trivial case that the "data" input is same as the detector level MC signal input. If there is large disagreement beyond statistical uncertainties after unfolding, this implies a problem in the method. Figure 7.10 shows the $p_T^{\mathrm{t,had}}$ and $\left|y^{\mathrm{t,had}}\right|$ distributions for the closure tests, demonstrating good agreement. A *stress* test checks that the unfolding is not biased by the truth distribution it is given. To do this, the detector level MC distribution is reweighted by some function, and unfolded using the nominal MC. Any non-closure between the particle level distribution reweighted by the same function and the result of the unfolding would imply that the unfolding is dependent upon the given particle level spectrum, which could hide effects of new physics in the real data. Two such tests are performed: first, the MC detector level "pseudo-data" is reweighted by

$$f\left(p_T^{\mathrm{hadtop}}, p_T^{\mathrm{leptop}}\right) = 1 + \frac{1}{400\,\mathrm{GeV}} \cdot \left(\frac{p_T^{\mathrm{hadtop}} + p_T^{\mathrm{leptop}}}{2} - 200\,\mathrm{GeV}\right), \tag{7.8}$$

where p_T^{hadtop} represents the particle level hadronic top quark p_T and similarly p_T^{leptop} the leptonic top p_T. The results of this test are shown in Figure 7.11. where it can be seen that there is good agreement between the stressed distributions. The nominal MC is also shown (in red) to demonstrate the size of the stress applied, which is large for the $p_T^{\mathrm{t,had}}$ distribution but minimal for the hadronic top quark $\left|y^{\mathrm{t,had}}\right|$. Thus, a second stress test is performed reweighting the pseudo-data using another equation, this time dependent on the $t\bar{t}$ system rapidity $y_{t\bar{t}}$

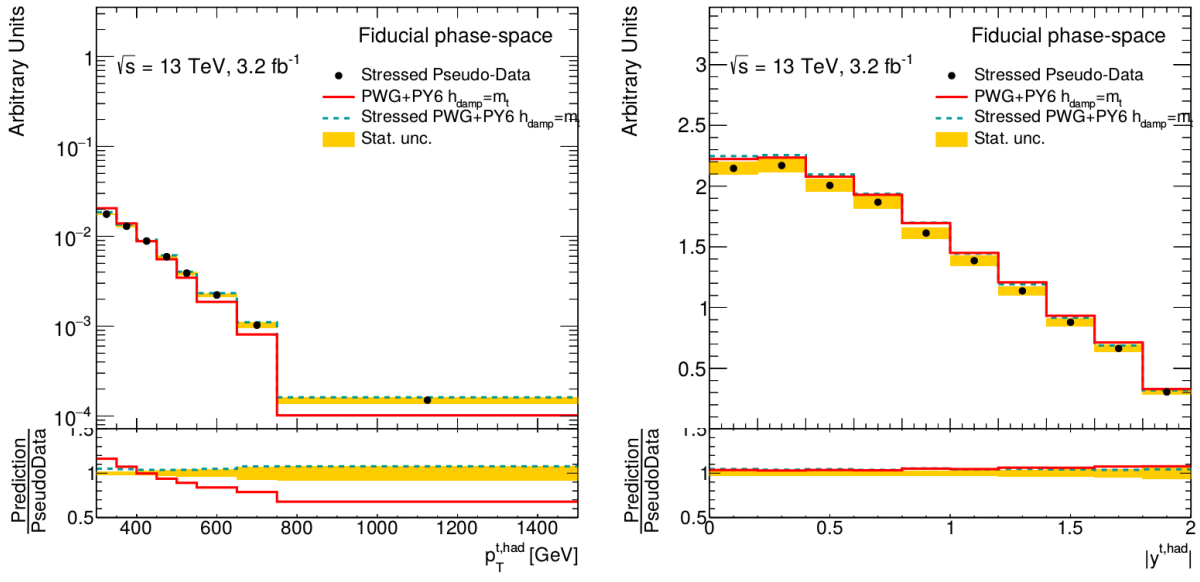
$$f\left(y_{t\bar{t}}\right) = 1 - 0.8 \cdot e^{-\frac{y_{t\bar{t}}^2}{0.04}}. \tag{7.9}$$

The results of this test are shown in Figure 7.12, where it is clear that despite the large stress applied to $\left|y^{\mathrm{t,had}}\right|$, the unfolding is able to correctly reproduce the stressed spectrum.



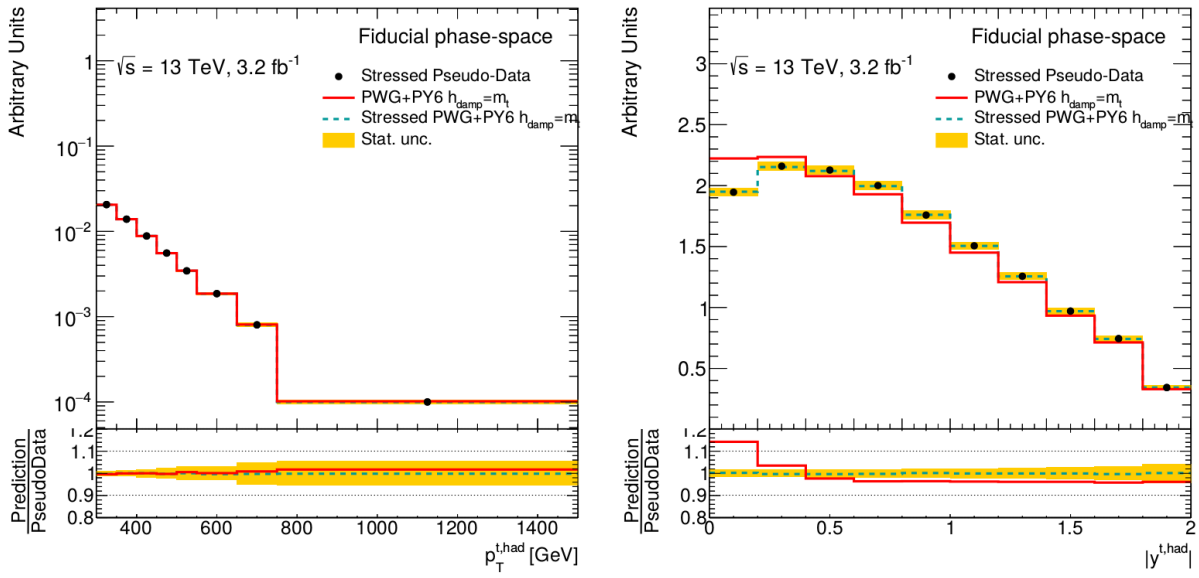
(a) (b)

Figure 7.10: The (a) $p_T^{\text{t,had}}$, and (b) $\left|y^{\text{t,had}}\right|$ closure tests, comparing the actual particle level distribution (labelled "prediction") with the unfolded detector level distribution (labelled "pseudo-data").



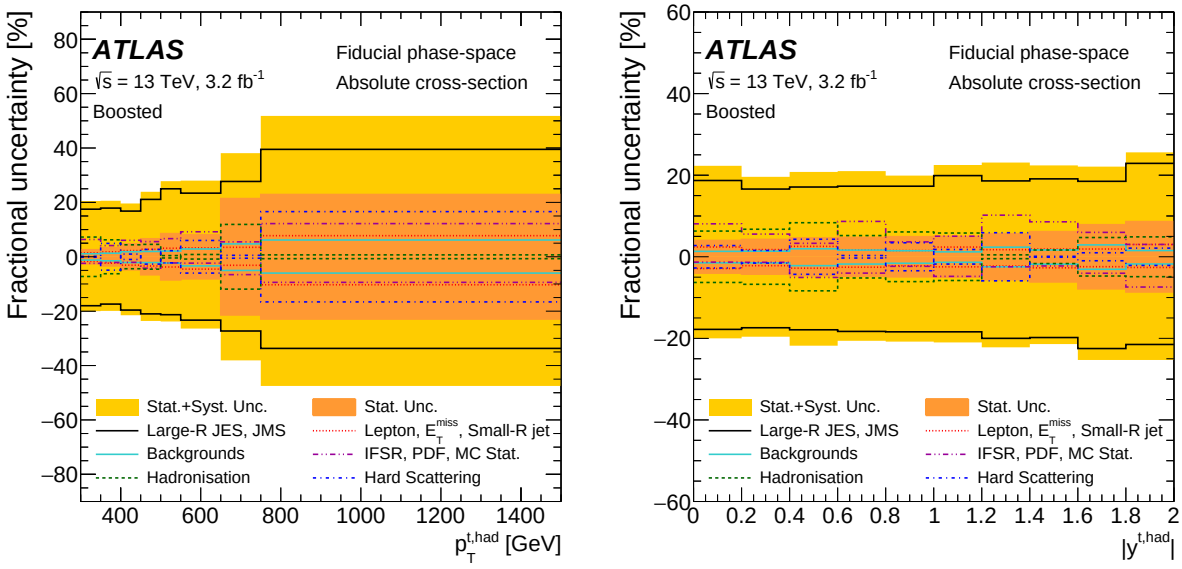
(a) (b)

Figure 7.11: The p_T dependent stress tests, comparing the nominal MC particle level distribution (in red) with the stressed detector level distribution after unfolding (black dots) and the equivalent stressed particle level distribution (dashed green), for (a) $p_T^{\text{t,had}}$, and (b) $\left|y^{\text{t,had}}\right|$.



(a) (b)

Figure 7.12: The $y^{t\bar{t}}$ dependent stress tests, comparing the nominal MC particle level distribution (in red) with the stressed detector level distribution after unfolding (black dots) and the equivalent stressed particle level distribution (dashed green), for (a) $p_T^{\text{t,had}}$, and (b) $\left|y^{\text{t,had}}\right|$.



(a) (b)

Figure 7.13: Plots showing the largest fractional contributions to the uncertainty of the absolute differential cross-section, after unfolding, for (a) the hadronic top p_T and (b) the hadronic top absolute rapidity. The largest contributions to the uncertainties come from the large jet energy scale (black line), the MC uncertainties for the matrix element generator, parton shower and hadronisation, and IFSR (dashed green line, dashed blue line, purple line respectively), and data statistics (orange band).

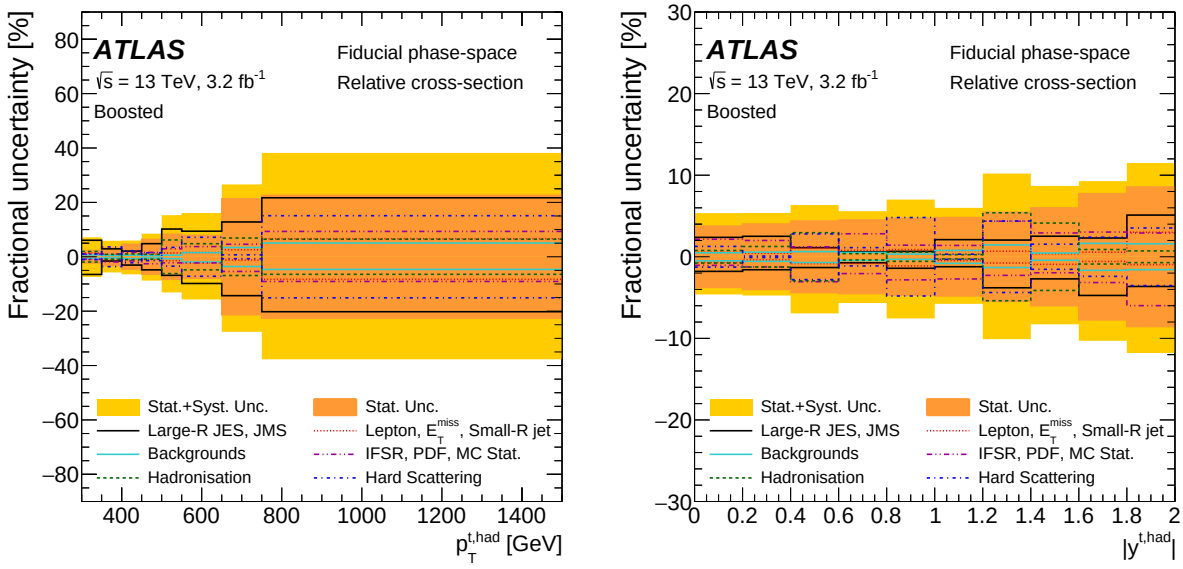
7.4 Systematic Uncertainties

All uncertainties on the objects used in the analysis, including the large jets, leptons, small jets, E_T^{miss}, and flavor tagging, are derived as discussed in Chapter 5. The uncertainty on the luminosity is ±2.1%. The MC modelling uncertainties are determined separately for the matrix element generator, parton shower and hadronisation model, initial & final state radiation, and PDF choice using dedicated samples:

- Matrix Element Generator: Unfolding of an aMC@NLO + Herwig++ sample is performed using corrections from a Powheg + Herwig++ sample, and the symmetrised difference is taken as uncertainty.

- Parton Shower and Hadronisation: The Powheg + Herwig++ sample is unfolded using the nominal Powheg + Pythia6 sample, and the symmetrised difference is taken as uncertainty.

- IFSR: Dedicated samples, where the factorisation and renormalisation scales in Pythia6 are varied simultaneously up or down by a factor of 2, are unfolded with the nominal Powheg + Pythia6 sample. In the down variation, the Powheg h_{damp} setting is increased by a factor of two, while in the up variation this is kept constant at $h_{\text{damp}} = m_{\text{top}}$, and



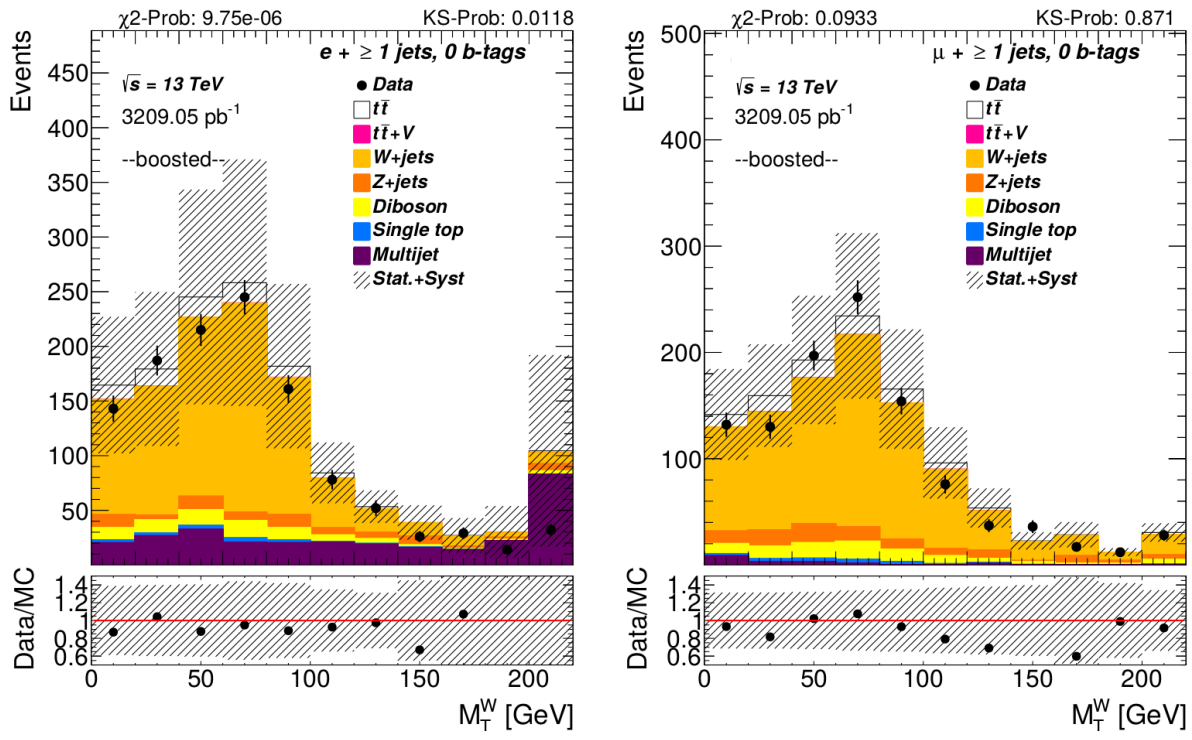
(a) (b)

Figure 7.14: Plots showing the largest fractional contributions to the uncertainty of the relative differential cross-section, after unfolding, for (a) the hadronic top p_T and (b) the hadronic top absolute rapidity. The largest contributions to the uncertainties come from the large jet energy scale (black line), the MC uncertainties for the matrix element generator, parton shower and hadronisation, and IFSR (dashed green line, dashed blue line, purple line respectively), and data statistics (orange band).

PYTHIA6 eigentunes radHi and radLo are also used. The envelope of these up and down variations are is taken as the uncertainty.

- PDF: Using the aMC@NLO + HERWIG++ sample, the central PDF4LHC15 PDF set is re-weighted to a set of 30 eigenvector PDFs, following the LHAPDF procedure [217]. The differences between each of these and the central PDF after unfolding are added in quadrature, as is the difference between the central and the CT10 PDF set, resulting in an uncertainty of $< 1\%$ in most bins.

The uncertainties on the MC-based background estimations are taken from the cross-section calculations they are normalised to, as discussed in Section 6.8. The data-driven corrections to the W+jets estimate require additional uncertainties, derived by varying one component at a time by its statistical uncertainty and rederiving the others. This is simple in the case of the charge asymmetry scale factors, while for the heavy flavor scale factors the additional constraint is applied that the total normalisation should be maintained. Uncertainties on the data-driven QCD multijet background are determined by comparing the results for the real and fake efficiency of two different control region definitions, with the difference symmetrised and taken as an uncertainty on the overall yield. Additionally, the normalisations of the residual backgrounds in the nominal control regions are varied by $\pm 20\%$ and the effect of this is propagated through the full procedure, with the difference taken as a further systematic uncertainty. Finally, a



(a) (b)

Figure 7.15: The m_T^W distributions in a 0 top-tag, 0 b-tag validation region, for (a) electron events and (b) muon events. The uncertainty bands show the systematics on the data driven background estimates, which are chosen to be conservative and are thus large in the control regions but have little effect on the final results.

mismodelling due to the multijet estimate is observed in the m_T^W distribution for electron events in a validation region in which there are 0 top-tagged large jets and 0 b-tagged small jets, shown in Figure 7.15. To account for this, a 100% uncertainty is applied on the multijet background for electron events in which $m_T^W > 150$ GeV. The addition of this uncertainty results in a sub-percent change in the integrated fiducial cross-section, with the largest effect being an additional 4% uncertainty on the highest $p_T^{\text{t,had}}$ bin. The major systematic uncertainties in the analysis relate to the jet energy scale of the large jet and the signal MC generation, as shown in Figure 7.13 for the absolute differential cross-section and in Figure 7.14 for the relative differential cross-section. Systematics are grouped together in these figures, with individual components added together in quadrature.

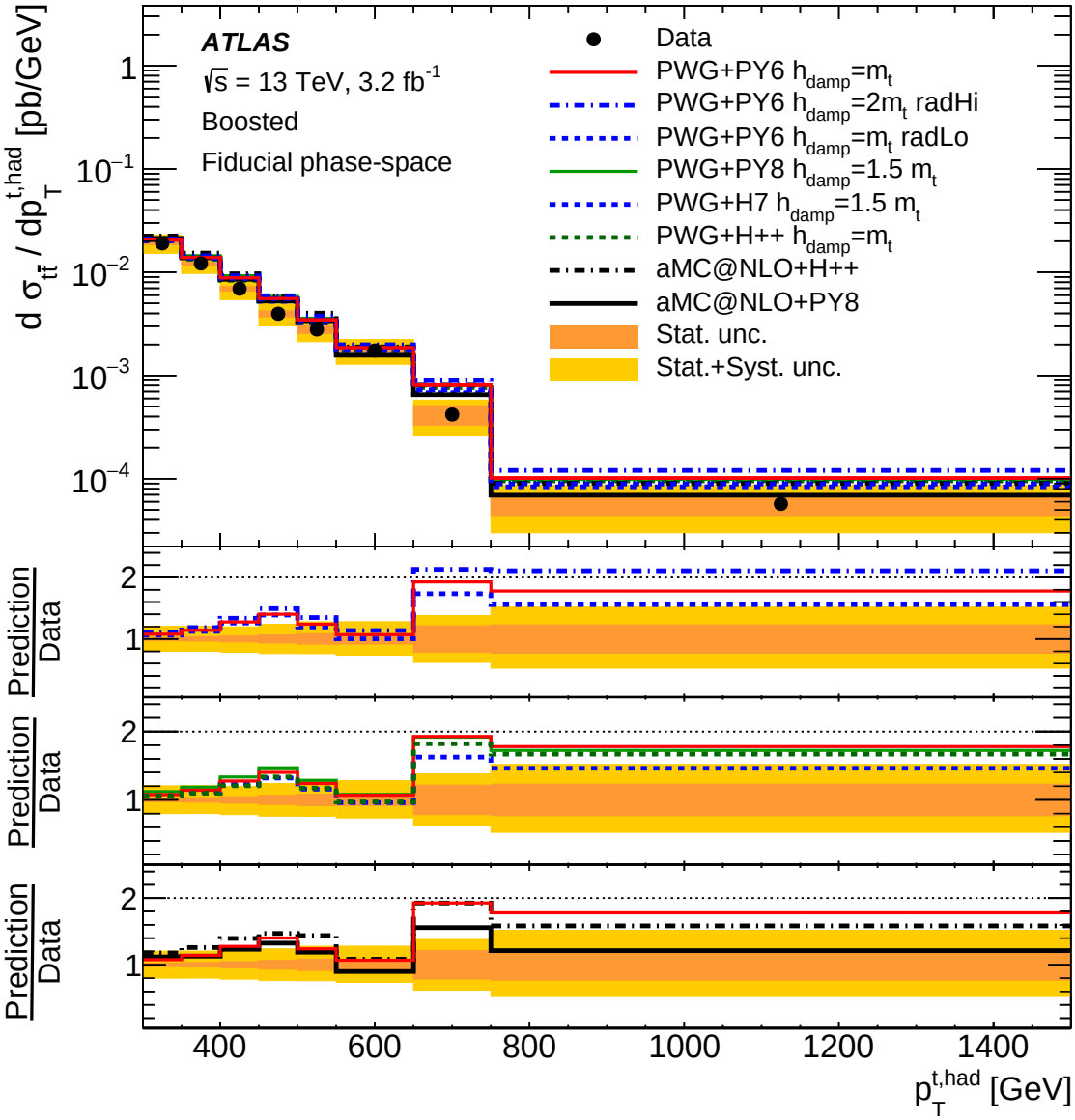


Figure 7.16: Normalised differential cross-sections in the boosted topology, as a function of top quark transverse momentum.

7.5 Results

The final distributions after unfolding are shown in Figures 7.16 and 7.17 (relative cross-section) and Figures 7.18 and 7.19 (absolute cross-section). Comparisons to a number of MC generators are included, with the lower ratio pad split in to three separate panels in order to make these more readable. The $p_T^{\mathrm{t,had}}$ distributions show a clear slope, with the MC generally predicting a harder spectrum than is observed. This mirrors what has been seen previously in Run 1, by both ATLAS [210, 211] and CMS [212, 213]. The $\left|y^{\mathrm{t,had}}\right|$ distributions, previously unmeasured in boosted events by ATLAS, show good agreement with all generators.

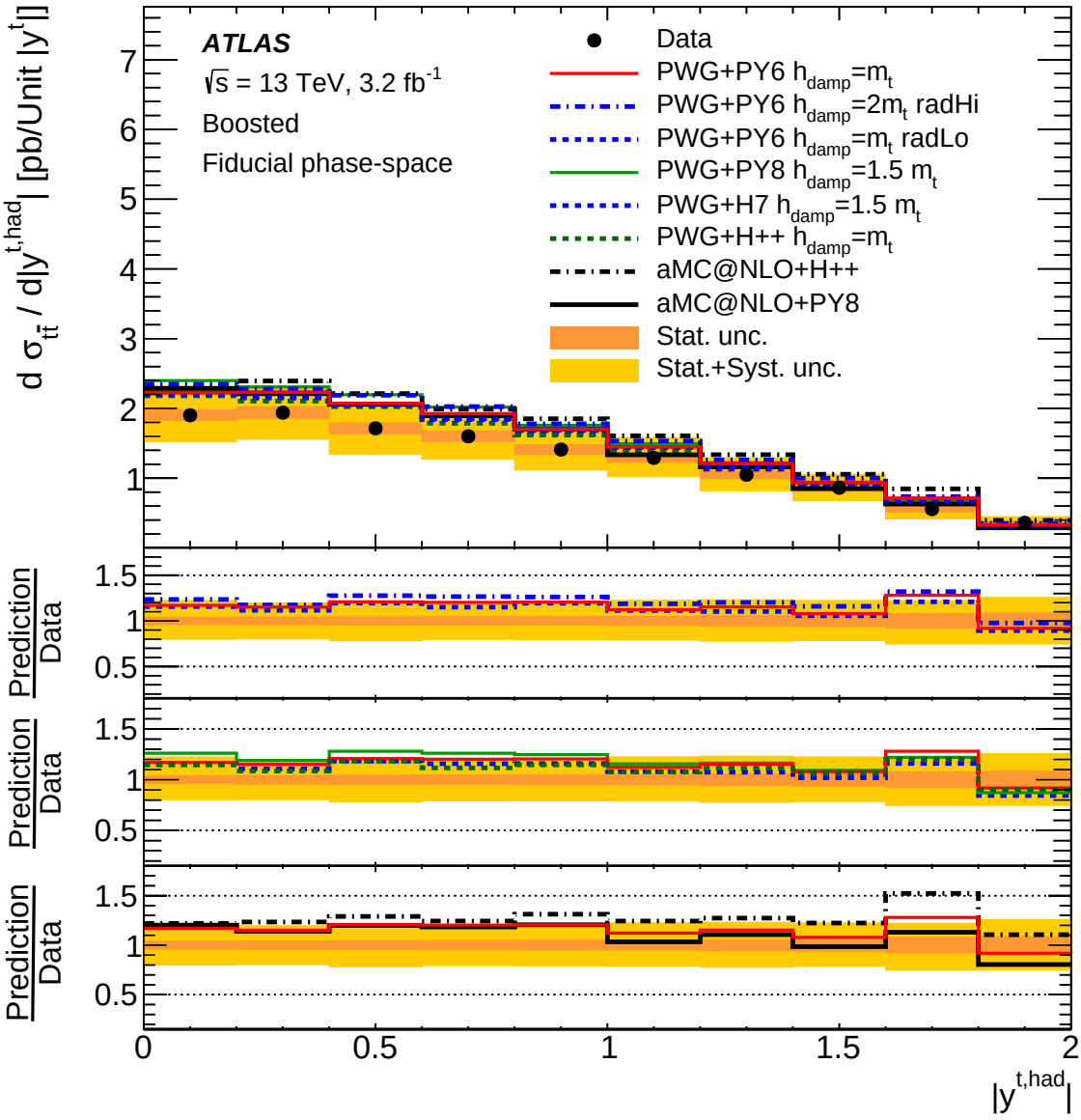


Figure 7.17: Normalised differential cross-sections in the boosted topology, as a function of top quark absolute rapidity.

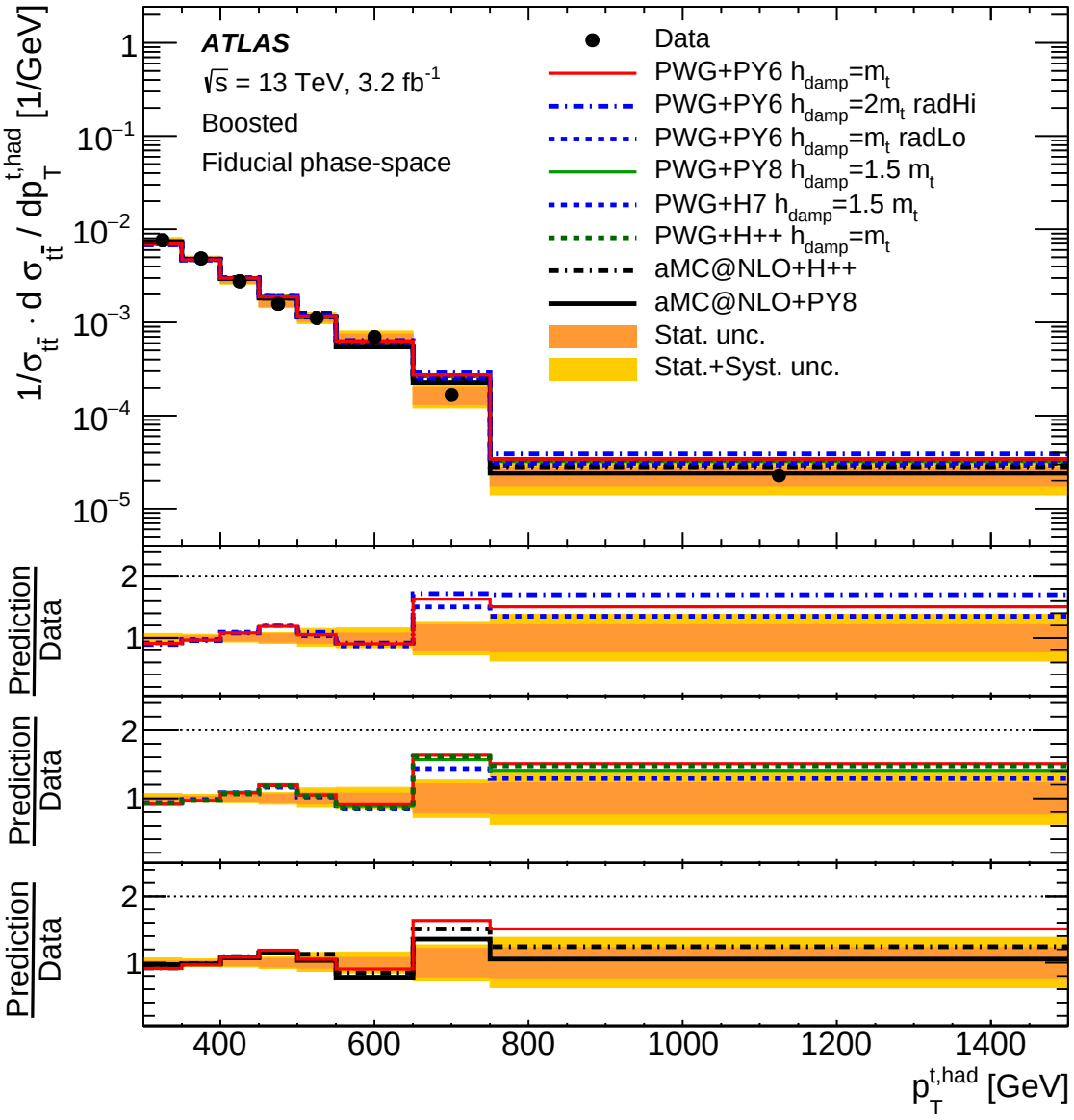


Figure 7.18: Absolute differential cross-sections in the boosted topology, as a function of top quark transverse momentum.

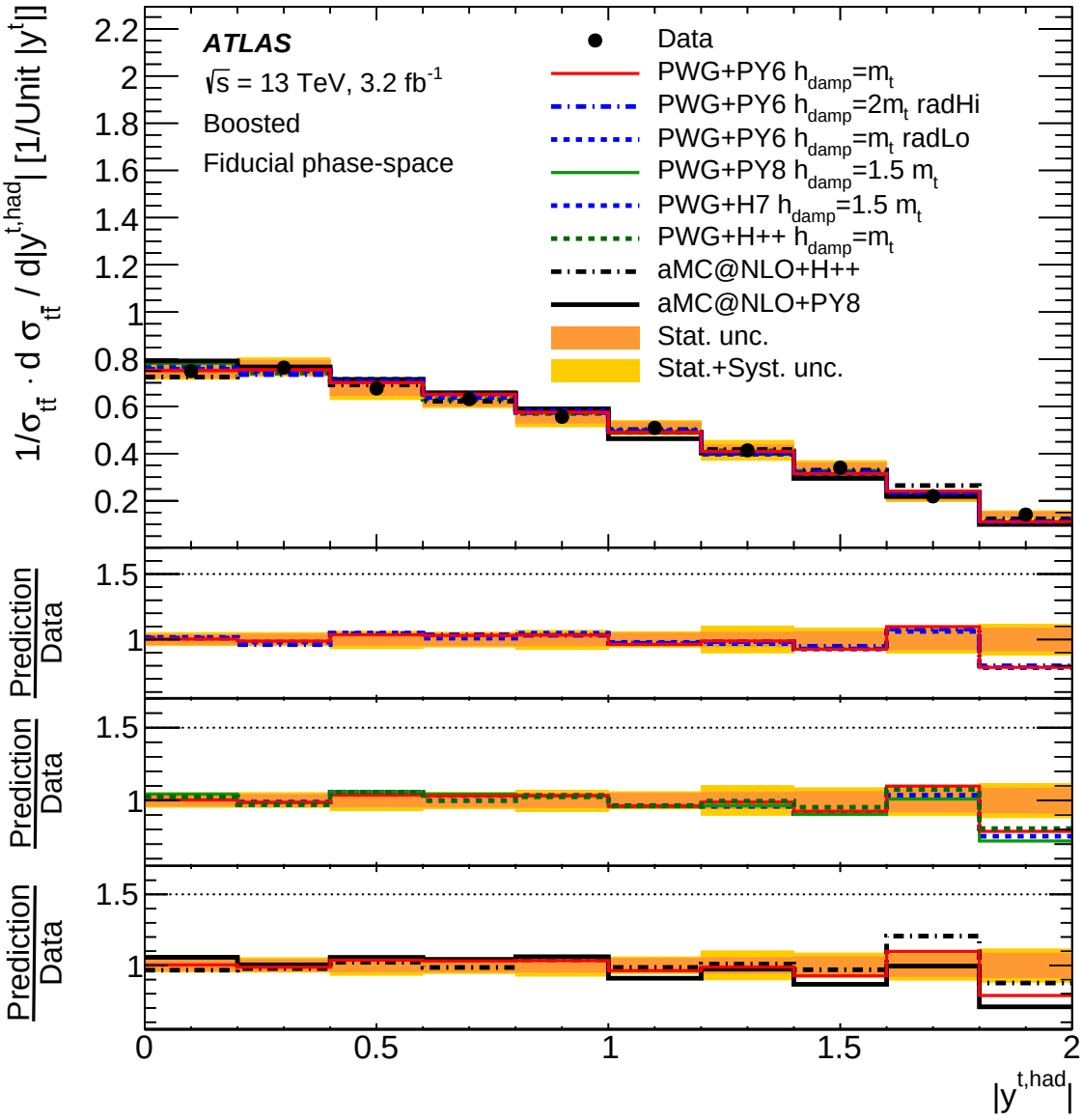


Figure 7.19: Absolute differential cross-sections in the boosted topology, as a function of top quark absolute rapidity.

Table 7.2: Comparisons of total fiducial cross-section in data and several MC generator set ups. All predictions are larger than the measured data by around 1σ.

Prediction	σ [pb]
POWHEG + PYTHIA6	2.95
aMC@NLO + HERWIG++	3.19
POWHEG + HERWIG++	2.84
POWHEG + PYTHIA8	3.01
POWHEG + PYTHIA6 radHi	3.10
POWHEG + PYTHIA6 radLo	2.89
Data	2.54 ± 18%

The total fiducial cross-section is compared to the same MC set ups, shown in Table 7.2, with the measurement around 1σ below each prediction. The POWHEG + HERWIG++ sample gives the prediction closest to the data, while the sample produced with aMC@NLO + HERWIG++ is the furthest out from the data.

Only a qualitative impression of agreement can be obtained by looking only at Figures 7.16 – 7.19, particularly as the bin-to-bin correlations are not shown in these plots. In order to quantify the degree of agreement between the data and the various generators a χ^2 value is computed. In order to do this, the covariance of the full statistical and systematic uncertainties between each pair of bins is evaluated. This is done using 10,000 pseudo-experiments, whereby the data is varied by a Poisson distribution in each bin, and then scaled by the relative variation of each systematic uncertainty. This shifted data is then unfolded using the nominal unfolding procedure, and the effect of these pseudo-experiments is then used to calculate the covariances. The signal modelling uncertainties, which are not well defined prior to unfolding, are evaluated in a slightly different manner, by assuming full correlation between each bin and simply multiplying the data by the relative uncertainty. The covariance matrices for each of the signal modelling uncertainties are then simply summed to the covariance from the detector systematics to produce the full $N_b \times N_b$ covariance matrix $\mathrm{Cov}_{\mathrm{N_b}}$, with N_b the number of bins.

The χ^2 for each model are then calculated, for the absolute distributions, with the equation

$$\chi^2_{\mathrm{abs}} = V^T_{N_b} \cdot \mathrm{Cov}^{-1}_{N_b} \cdot V_{N_b}, \tag{7.10}$$

where V_{N_B} is the vector of differences between data and prediction. In the case of the normalised distributions, the normalisation constraint on the relative differential cross section removes a degree of freedom and lowers the rank of the matrix to $N_b - 1$

$$\chi^2_{\mathrm{norm}} = V^T_{N_b-1} \cdot \mathrm{Cov}^{-1}_{N_b-1} \cdot V_{N_b-1}, \tag{7.11}$$

where V_{N_b-1} is then the vector of differences between data and prediction with one element

Table 7.3: Summary of χ^2 and p-values for the absolute distributions. The best description of the $p_T^{\mathrm{t,had}}$ distribution is with Powheg+Herwig7, and the best description of $\left|y^{\mathrm{t,had}}\right|$ is from Powheg+Herwig++.

	$p_{\mathrm{T}}^{t,\mathrm{had}}$		$\lvert y^{t,\mathrm{had}}\rvert$	
	χ^2 / NDF	p-val	χ^2 / NDF	p-val
Powheg+Pythia6	14.7 / 8	0.06	11.0 / 10	0.36
Powheg+Pythia6 (radHi)	19.5 / 8	0.01	12.3 / 10	0.27
Powheg+Pythia6 (radLo)	15.0 / 8	0.06	10.0 / 10	0.44
MadGraph5_aMC@NLO+Herwig++	17.9 / 8	0.02	12.8 / 10	0.24
Powheg+Herwig++	14.1 / 8	0.08	8.0 / 10	0.63
MadGraph5_aMC@NLO+Pythia8	12.8 / 8	0.12	20.4 / 10	0.03
Powheg+Pythia8	16.7 / 8	0.03	18.4 / 10	0.05
Powheg+Herwig7	11.9 / 8	0.15	11.7 / 10	0.30

Table 7.4: Summary of χ^2 and p-values for the relative distributions. The best description of the $p_T^{\mathrm{t,had}}$ distribution is with Powheg+Herwig7, and the best description of $\left|y^{\mathrm{t,had}}\right|$ is from Powheg+Herwig++.

	$p_{\mathrm{T}}^{t,\mathrm{had}}$		$\lvert y^{t,\mathrm{had}}\rvert$	
	χ^2 / NDF	p-val	χ^2 / NDF	p-val
Powheg+Pythia6	10.2 / 7	0.18	2.9 / 9	0.97
Powheg+Pythia6 (radHi)	11.3 / 7	0.12	2.9 / 9	0.97
Powheg+Pythia6 (radLo)	11.5 / 7	0.12	2.8 / 9	0.97
MadGraph5_aMC@NLO+Herwig++	11.1 / 7	0.13	4.6 / 9	0.87
Powheg+Herwig++	10.7 / 7	0.15	2.5 / 9	0.98
MadGraph5_aMC@NLO+Pythia8	10.9 / 7	0.14	7.2 / 9	0.62
Powheg+Pythia8	11.3 / 7	0.13	4.3 / 9	0.89
Powheg+Herwig7	9.9 / 7	0.20	3.6 / 9	0.94

discarded[1], and $\mathrm{Cov}^{-1}_{N_b-1}$ is the corresponding inverted and reduced covariance matrix. The χ^2, and corresponding p-values, for each generator set up are shown in Table 7.3 for the absolute and Table 7.4 for the normalised distributions. From these, it is shown that the best description of the $p_T^{\mathrm{t,had}}$ spectrum comes from Powheg+Herwig7, while Powheg+Herwig++ is best for $\left|y^{\mathrm{t,had}}\right|$.

The correlation matrix for the p_T distribution is shown in Figure 7.5, where it can be seen that the measurements of neighbouring bins are highly correlated, while there is often corresponding anti-correlation for non-neighbouring bins due to the normalisation constraint. These strong correlations display at a glance how difficult it can be to interpret the final results by eye, and underline the importance of the χ^2 and p-values.

[1]The choice of which element to discard is arbitrary, and does not affect the calculation.

Table 7.5: Correlation matrix for the normalised p_T distribution.

Bin GeV	300–350	350–400	400–450	450–500	500–550	550–650	650–750	750–1500
300–350	**1.00**	0.36	−0.42	−0.57	−0.46	−0.47	−0.53	−0.52
350–400	0.36	**1.00**	−0.01	−0.22	−0.03	0.04	−0.23	−0.11
400–450	−0.42	−0.01	**1.00**	0.34	0.30	0.50	0.27	0.37
450–500	−0.57	−0.22	0.34	**1.00**	0.51	0.45	0.48	0.49
500–550	−0.46	−0.03	0.30	0.51	**1.00**	0.59	0.44	0.51
550–650	−0.47	0.04	0.50	0.45	0.59	**1.00**	0.43	0.54
650–750	−0.53	−0.23	0.27	0.48	0.44	0.43	**1.00**	0.44
750–1500	−0.52	−0.11	0.37	0.49	0.51	0.54	0.44	**1.00**

7.5.1 Conclusions

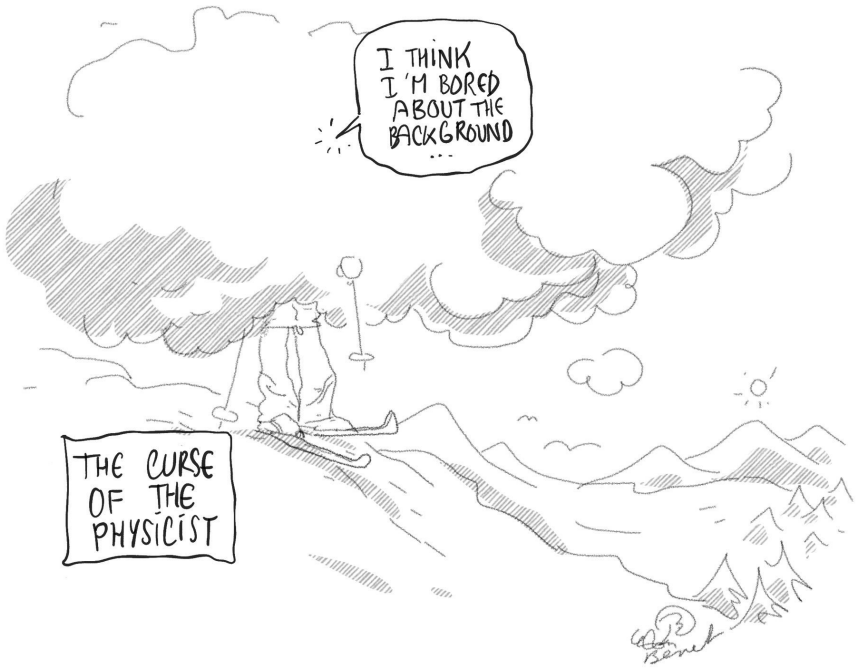
The results presented in this chapter represent the first measurement of boosted top quark kinematics at $\sqrt{s}$ = 13 TeV and were published in JHEP in August 2017 [1]. Further, full uncertainties breakdowns, correlations, and covariance matrices for each distribution have been released in HepData [218], allowing the results to be reused and reinterpreted by the high energy physics community.

The precision of the analysis is largely limited by the uncertainties on large jets, which have since been reduced using in situ techniques and the increased data set now available [219]. Alternative techniques for the boosted regime, such as the use of reclustered jets (as used in the analysis presented in Chapter 8) also allow other ways around these large systematics. Other major uncertainties arise from the MC generator systematics, which involve comparison of different generator set ups. Recently, techniques have been developed to derive these uncertainties using weights within a single generator set up, which has the potential to reduce these uncertainties by better factorising the various effects and comparing directly the same events with different settings. Additionally, the tuning of these MC set ups can be improved as more measurements are made, and indeed the results presented have already been used to do this [220], resulting in lower modelling systematics in future measurements. Finally, the granularity of the binning used, as well as the p_T reach possible, will benefit from the 140 fb^{-1} of data that is now available from the full Run 2 of the LHC.

While the $\left|y^{\mathrm{t,had}}\right|$ spectrum shows good agreement with the Standard Model, the $p_T^{\mathrm{t,had}}$ distribution shows some tension with the SM expectation, and this disagreement is largest at high p_T. There are hints that this can be explained by NNLO QCD [221] and NLO electroweak [222] corrections, though no appropriate predictions were available for direct comparison at the time of publication. Another explanation could be that the differences between data and prediction come from some residual effect of BSM physics, and the results presented are potentially useful in effective field theory fits that search for such BSM effects [223–225].

Chapter 8

Searching for $t\bar{t}H, H \to b\bar{b}$ using Boosted Objects



– @lisbouche

One of the primary goals of the ATLAS Collaboration during Run 2 of the LHC was the discovery of the $t\bar{t}H$ process. This chapter presents the search carried out in the $t\bar{t}H, H \to b\bar{b}$ channel, using 36 fb^{-1} of $\sqrt{s}$ = 13 TeV data collected during 2015 and 2016, and published in [2]. Particular focus is given to the analysis in the lepton+jets boosted regime, in which both the hadronically decaying top quark and the Higgs boson are produced at high p_T, which was not published during Run 1.

8.1 Event Selection and Categorisation

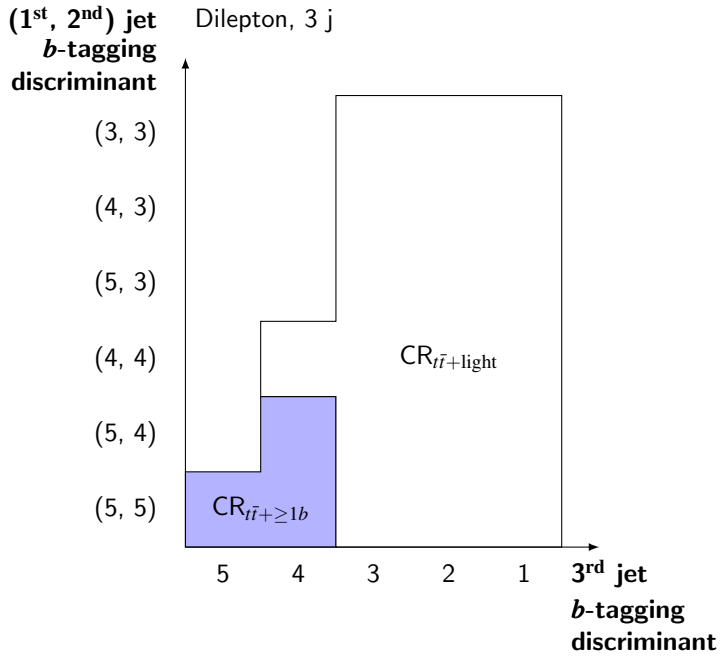
Events are selected targeting the lepton+jets boosted, lepton+jets resolved, and dilepton resolved topologies, based on the presence of objects as described in Chapter 5. All are required to have a lepton, of $p_T > 27$ GeV and $|\eta| < 2.5$, which passes single lepton (e / μ) triggers as described in Section 4.2.5. In each category, there must also be at least two small jets b-tagged at the 85% working point. Events are subsequently categorised into signal and control regions based on multiplicity and pseudo-continuous b-tagging (PCB) of the small jets, whereby each jet is given a score based on the tightest working point at which it is b-tagged: 60%=5, 70%=4, 77=3, 85=2, 100%=1, and the jets are then ordered based on this score. PCB is further described in Section 5.3.2. Signal regions are chosen which have a high signal yield, while control regions are chosen to help constrain each of the three primary backgrounds of $t\bar{t}+ \geq 1b$, $t\bar{t}+ \geq 1c$, and $t\bar{t}$+light. Events containing ≥ 2 reconstructed τ leptons are vetoed from all regions to ensure orthogonality with the $t\bar{t}H$ analysis in the multilepton final state [226], with negligible effect on any results.

8.1.1 Resolved Dilepton

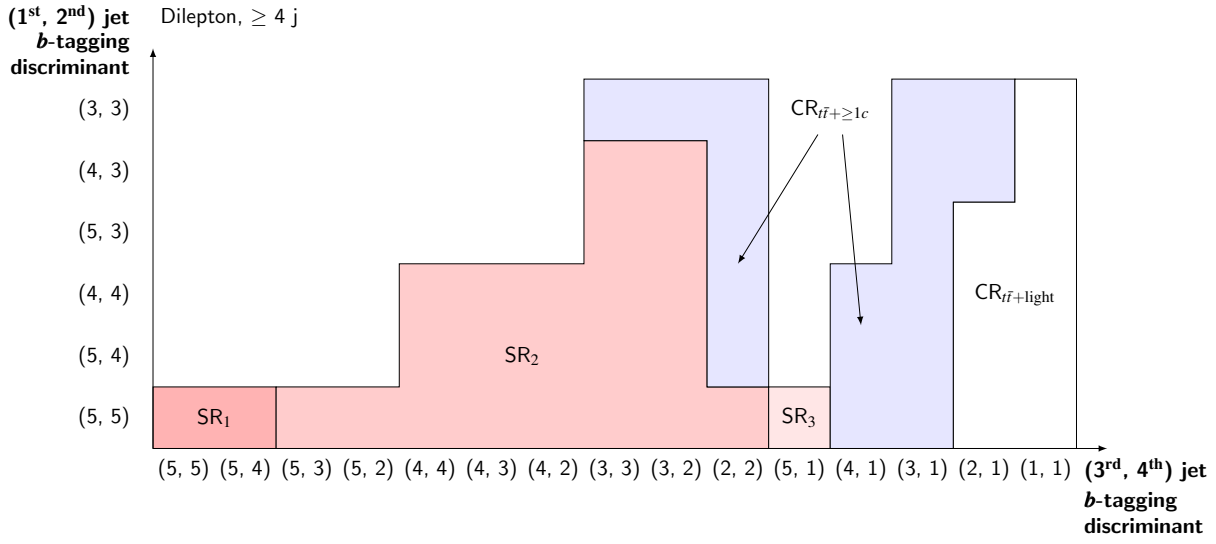
Resolved dilepton events are required to have ≥ 3 small jets and an additional, oppositely charged lepton with $p_T > 10$ GeV. If both leptons are electrons, this p_T requirement is tightened to $p_T > 15$ GeV in order to reduce the number of events resulting from fake leptons, and in the ee or $\mu\mu$ channels the invariant mass of the dilepton system must be greater than 15 GeV and outside the Z-boson mass window 83 – 99 GeV. Dilepton events are then further categorised into events with exactly 3 or ≥ 4 small jets, and further by PCB. The final categorisation is summarised in Figure 8.1.

8.1.2 Resolved Lepton+Jets

For the resolved lepton+jets events, there must be ≥ 5 small jets and no second lepton with $p_T > 10$ GeV. Events are split into categories with exactly 5 or ≥ 6 small jets, and subsequently split into regions based on PCB, analogously to the dilepton case. The exact definitions of these regions is summarised in Figure 8.2.

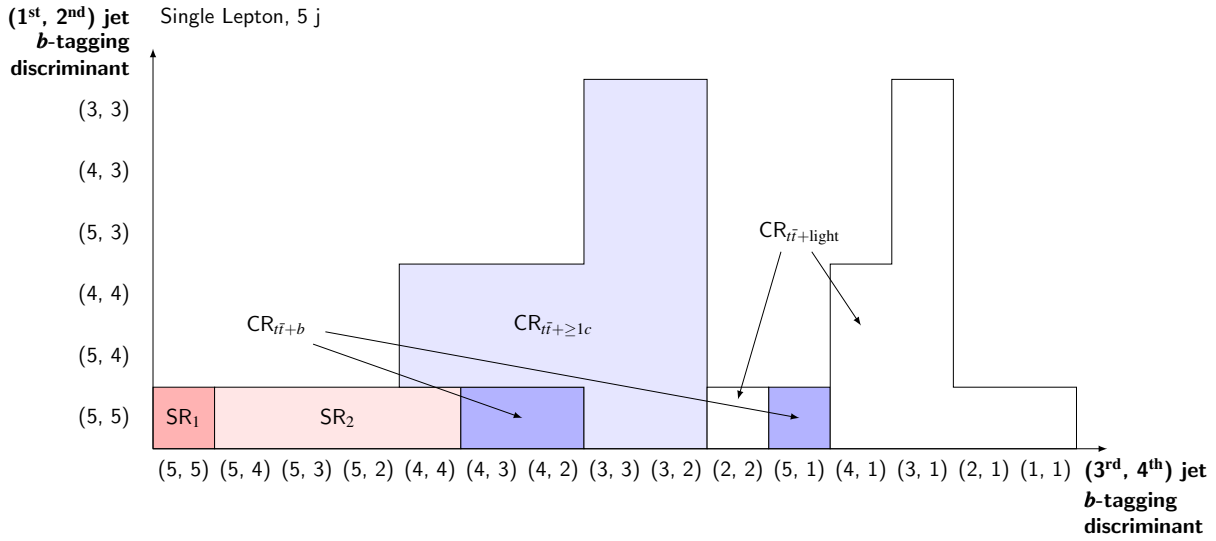


(a)

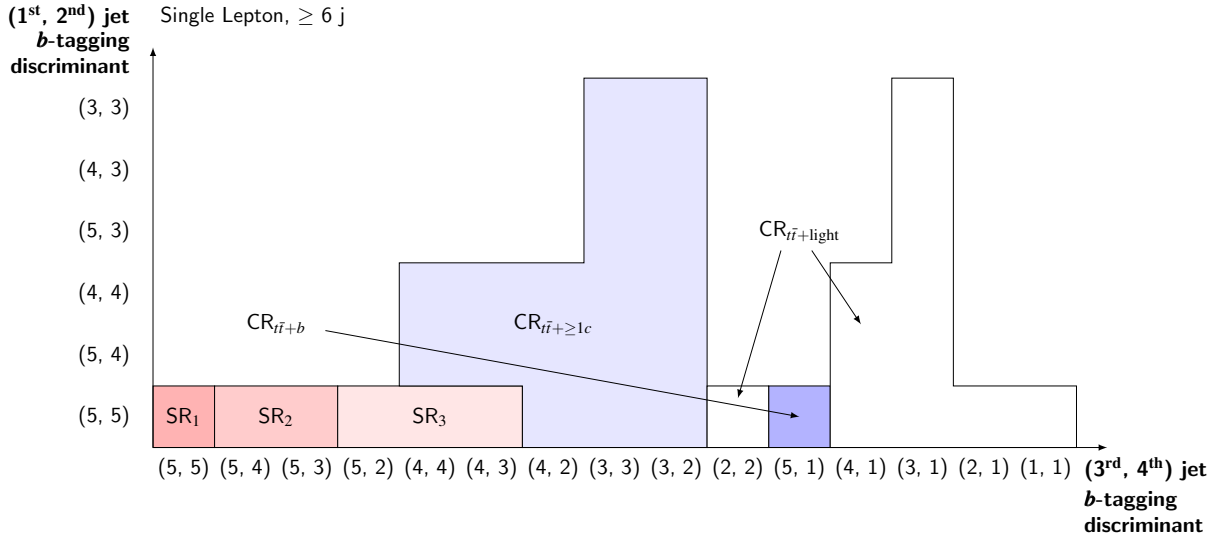


(b)

Figure 8.1: Event Categorisation based on PCB scores in the dilepton channel, for events with (a) 3 small jets, and (b) ≥ 4 small jets. The y-axis represents the PCB bin for the leading and sub-leading b-jet, and the x-axis represents the PCB bin for the (a) third, and (b) third and fourth b-jet.



(a)



(b)

Figure 8.2: Event Categorisation based on PCB scores in the resolved lepton+jets channel, for events with (a) 5 small jets, and (b) $\geq$ 6 small jets. The y-axis represents the PCB bin for the leading and sub-leading b-jet, and the x-axis represents the PCB bin for the third and fourth b-jet.

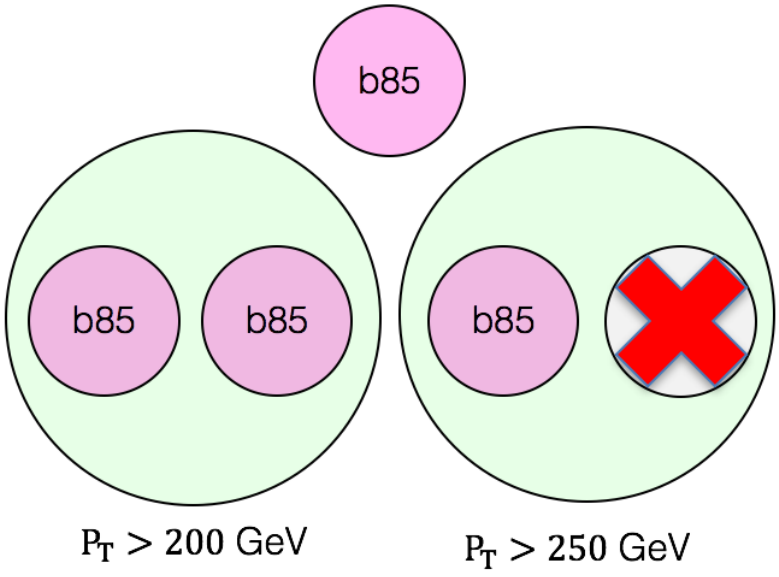
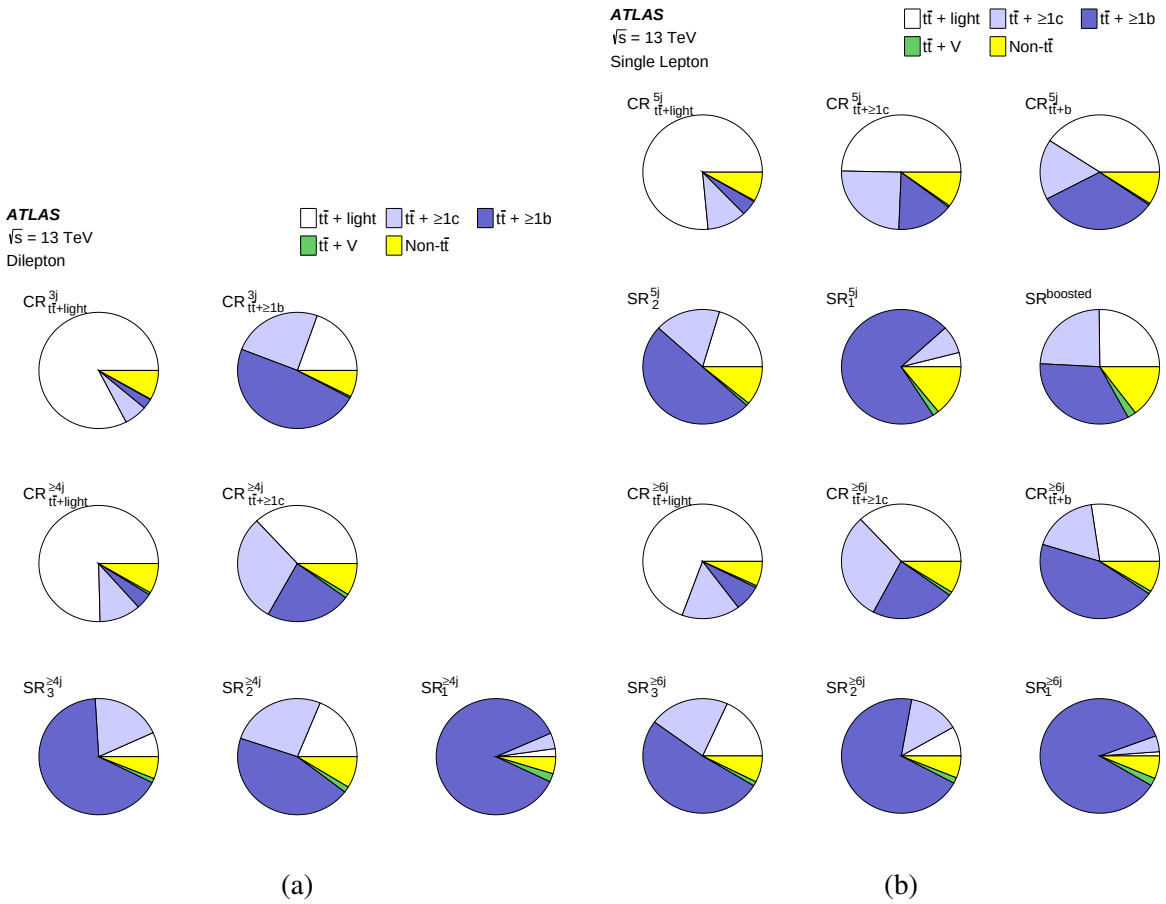


Figure 8.3: Cartoon summary of the boosted selection in the $t\bar{t}H, H \to b\bar{b}$ analysis. The large green circles represent reclustered jets while the smaller pink circles labelled "b85" represent small jets b-tagged at the 85% WP. The small white circle with the red cross represents a small jet that does not pass the b-tagging requirement. The location either inside or outside the reclustered jets represents which small jets are subjets and which are additional jets.

8.1.3 Boosted Lepton+Jets

Finally, a boosted lepton+jets region is defined, which must contain a single lepton and at least two reclustered jets with $p_T > 200$ GeV. Small jets which are clustered into these are referred to as *subjets*. Further requirements are placed on the b-tagging of the subjets, all using the 85% working point. One of the reclustered jets, referred to as the Higgs candidate, must contain at least two b-tagged subjets, matching the expected signature of a $H \to b\bar{b}$ decay. A second reclustered jet must pass $p_T > 250$ GeV, and contain exactly one b-tagged subjet and at least one subjet that has not been b-tagged – this reclustered jet is referred to as the hadronic top candidate. The requirement of ≥ 1 light jet, instead of the ≥ 2 expected from the intermediate W boson, enhances the selection efficiency at high p_T where the W decay products become collimated and are reconstructed as a single jet. Finally, there must be at least one *additional* jet – defined as small jets that are not reclustered into either the top or Higgs candidates – at least one of which must be b-tagged, consistent with the b-jet resulting from a leptonic top decay. This event selection in the boosted channel is summarised in Figure 8.3. Events which pass the boosted selection are vetoed from the resolved regions in order to ensure orthogonality, with an expected signal yield of 16.9 events in the boosted region, of which 2.6 are vetoed from the tightest resolved signal region.



(a) (b)

Figure 8.4: The relative background fractions in each region, for (a) dilepton events, and (b) lepton+jets events. Control regions are chosen such that they are enriched in different components of the $t\bar{t}$ background, in order to constrain these in the fit.

8.2 Background Modelling

In $t\bar{t}H, H \to b\bar{b}$, the primary background comes from $t\bar{t}$+jets production, with only small contributions from $t\bar{t}V$ and non-$t\bar{t}$ backgrounds, as shown in Figure 8.4. The $t\bar{t}$ background is further split into three components, depending on the heavy flavor composition of the additional jets. The production of $t\bar{t}+\geq 1b$ is of particular importance as it can have an identical final state to the $t\bar{t}H, H \to b\bar{b}$ signal. Sizeable contributions from $t\bar{t}+\geq 1c$ and $t\bar{t}$+light jets are also present, particularly in the looser signal regions. Additionally, the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ backgrounds are particularly difficult to model, as in most MC setups the additional heavy flavor jets are calculated only in the parton shower. Control regions are thus included in the fit in order to get a handle on these backgrounds, as shown in Figures 8.1 and 8.2. Nominally, the Powheg + Pythia8 MC samples are used for these backgrounds, with the overall normalisations of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ components left freely floating in the fit, represented by two nuisance parameters k_b and k_c. For the $t\bar{t}+\geq 1b$ background, further sub components are defined: those events with one additional b-quark matched to a small jet are labelled $t\bar{t}$+b, two additional b-jets are called

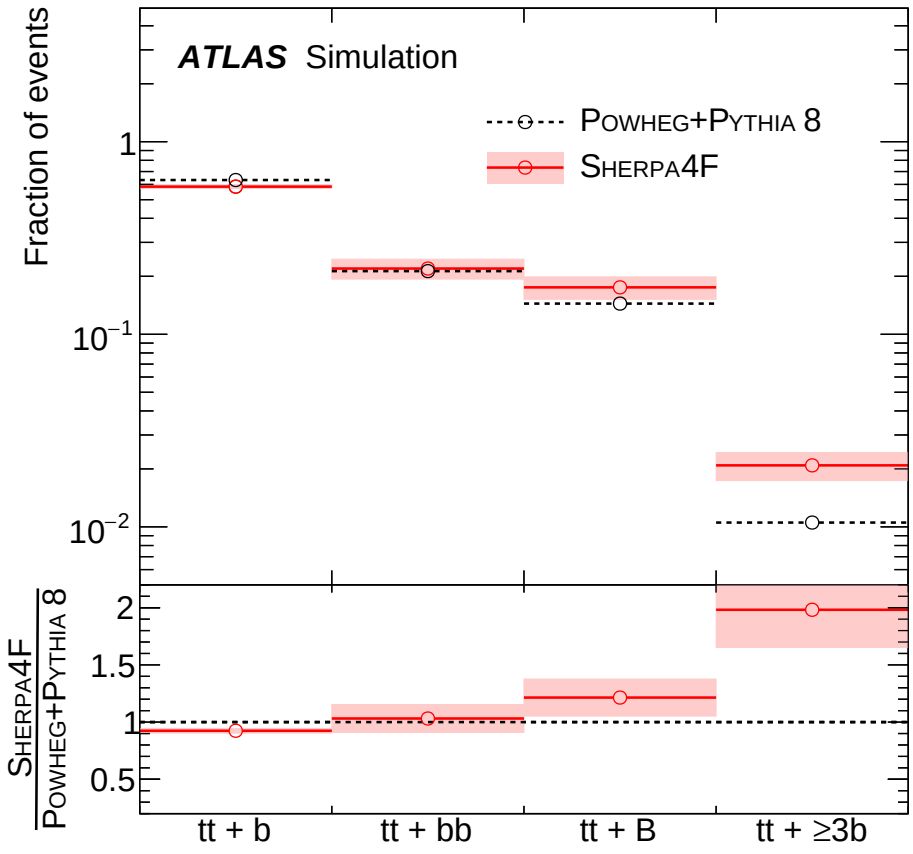
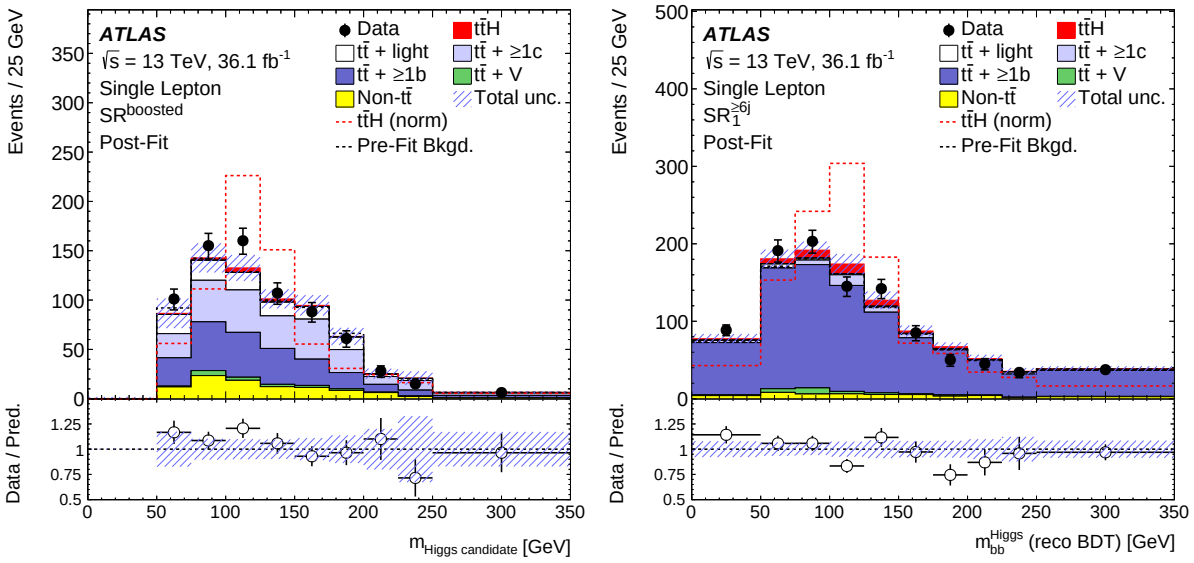


Figure 8.5: The relative fractions of the various $t\bar{t}$+b sub-components, for the nominal POWHEG + PYTHIA8 sample and the SHERPA4F sample that the others are reweighted to, showing that the fractions are not in good agreement between samples.

$t\bar{t}+b\bar{b}$, those with three or more additional b-jets are referred to as $t\bar{t}+\geq$ 3b, and finally those with a single additional jet which contains more than one b-quark are labelled $t\bar{t}$+B. The relative fraction of these is calculated inclusively for all regions using the SHERPA4F $t\bar{t} + b\bar{b}$ MC sample, and nominal predictions, as well as the alternate $t\bar{t}$ samples used for systematics, are reweighted to match these fractions. These fractions are shown for the POWHEG + PYTHIA8 and SHERPA4F samples in Figure 8.5, where it can be seen that these fractions are not in agreement within the MC statistics of the available samples, motivating the reweighting procedure to the higher precision SHERPA prediction.

The smaller backgrounds are all taken directly from MC, with the exception of the QCD multijet background. In the dilepton channel, this is calculated from MC with a normalisation correction, derived in events in which both leptons have the same electromagnetic charge. In the lepton+jets channel, a matrix method, similar to that described in Section 7.2, is used, though this background is negligible in the boosted channel as well as the tightest resolved SR. The full details of all of the MC samples are provided in Section 6.8.



(a) (b)

Figure 8.6: The Higgs candidate mass distributions in (a) the boosted signal region, and (b) the tightest resolved signal region. The signal $t\bar{t}H$ MC is shown in red, with the dashed line normalised to the background, and the solid red line normalised to the SM expectation and stacked on top of the backgrounds.

8.3 Multivariate Strategy

In the resolved channels, event reconstruction is a complex process involving multiple MVA techniques, including a reconstruction BDT, Likelihood discriminant, and Matrix Element Method. The full details of these can be found in [2]. One of the motivating factors for the boosted channel is the greatly simplified event reconstruction, and as such, no intermediate MVAs are required. This is illustrated by the reconstruction efficiency for the Higgs boson, defined as the fraction of selected events in signal MC in which both of the Higgs b-quark daughters are matched to the Higgs candidate. The resolved lepton+jets reconstruction BDT, which tests each possible combination of small jet assignments to the decay products of the hadronic top quark, leptonic top quark, and Higgs boson, and outputs a score indicating the combination most likely to be correctly assigned, achieves 48% efficiency in the purest resolved signal region. In the boosted channel, with no such reconstruction MVA, 47% of the selected reclustered jets are correctly matched. The Higgs candidate mass distributions for these two regions are shown in Figure 8.6.

A classification BDT is trained independently in each signal region to separate signal $t\bar{t}H$ events from the backgrounds, using the TMVA framework [74]. The input variables to the BDT in the boosted region, summarised in Table 8.1, were chosen from a longer list of well-modelled variables with good signal/background separation by iteratively removing the lowest performing variable until a significant drop in performance was observed. The hyperparameters, listed in Table 8.2, were optimised by hand to maximise performance and minimise overtraining. The

Table 8.1: Input variables for the classification BDT in the boosted signal region. The *additional* b-jet is the leading b-tagged jet that is not reclustered into the top or Higgs jets.

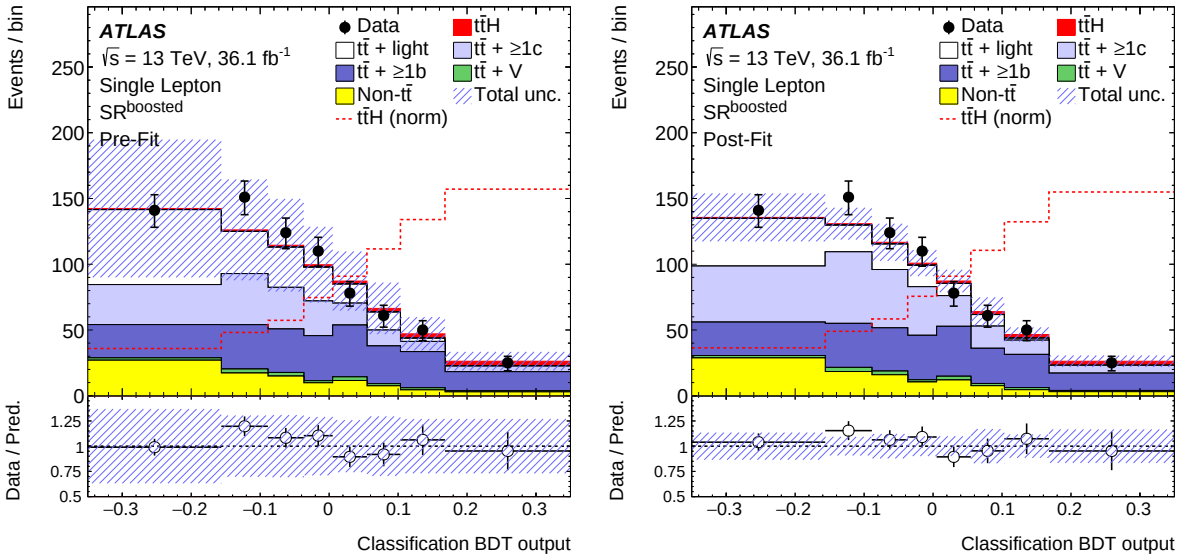
Variable	Definition
$\Delta R_{H,t}$	ΔR between the Higgs boson and top quark candidates
$\Delta R_{t,b^{\text{add}}}$	ΔR between the top quark candidate and additional b-jet
$\Delta R_{H,b^{\text{add}}}$	ΔR between the Higgs boson candidate and additional b-jet
$\Delta R_{H,\ell}$	ΔR between the Higgs boson candidate and lepton
m_{Higgs}	Higgs boson candidate mass
$\sqrt{d_{12}}_{\text{top}}$	Top quark candidate first splitting scale
$w_{b\text{-tag}}$	Sum of b-tagging discriminants of all b-jets
$w^{\text{add}}_{b\text{-tag}}/w_{b\text{-tag}}$	Ratio of sum of b-tagging discriminants of additional b-jets to all b-jets

Table 8.2: TMVA BDT hyperparameters used for the classification BDT in the boosted region.

Hyperparameter	Value
Number of Trees	700
Maximum Depth	3
Minimum Node Size	2%
AdaBoost β	1.0
Bagging Fraction	50%

output distribution of this BDT is shown in Figure 8.7, with similar distributions built for the resolved signal regions.

A profile likelihood fit (described in Section 3.3) is performed simultaneously on all of the signal and control regions, as defined in Section 8.1. In the signal regions, the classification BDT distributions are used in the fit. In the $t\bar{t}$+≥1c control regions, the scalar sum of the p_T of all of the small jets, H_T^{had}, is used, while all other control regions use a single bin.



(a) (b)

Figure 8.7: The BDT discriminant distribution in the boosted signal region, (a) with the nominal background predictions and systematic uncertainties, and (b) after the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ normalisations and the systematic uncertainties have been constrained by the profile likelihood fit. The signal $t\bar{t}H$ MC is shown in red, with the dashed line normalised to the background, and the solid red line normalised to the SM expectation and stacked on top of the backgrounds.

8.4 Systematic Uncertainties

A total of 245 systematic uncertainties are considered, summarised in Table 8.3. If the overall effect of the uncertainty is smaller than 1% it is considered negligible and removed (*pruned*) from the fit. Those which apply to multiple processes are applied to each process and pruned separately. Systematic uncertainties on the detector performance and object reconstruction follow those described in Chapter 5, and the uncertainty on the luminosity of the combined 2015+2016 dataset is 2.1%.

The major sources of uncertainty in the analysis are related to the modelling of the various processes. The signal modelling has systematic uncertainties of $^{+5.8\%}_{-9.2\%}$ due to scale, ±3.6% due to PDF, and ±2.2% due to branching fraction [6, 198–202]. A further uncertainty due to the choice of parton shower is applied as described in Section 6.8. There are a total of 23 different systematic components on the $t\bar{t}$+jets backgrounds, summarised in Table 8.4. In particular, the overall normalisation of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ backgrounds are left freely floating in the fit, with no prior assigned. This reflects the poorly understood modelling of these components, and lack of good corresponding measurement of these processes at the time of publication.

Table 8.3: Breakdown of the sources of systematic uncertainty in the $t\bar{t}H, H \to b\bar{b}$ analysis. The second column, labelled "Type", represents whether the uncertainty is on shape (S), normalisation (N), or both (SN). The third column, labelled "Comp", represents the number of components for each grouping.

Systematic Uncertainty	Type	Comp.
Experimental Uncertainties		
Luminosity	N	1
Pileup Modelling	SN	1
Physics Objects		
Electron	SN	6
Muon	SN	15
Taus	SN	3
Jet energy scale	SN	20
Jet energy resolution	SN	2
Jet vertex tagger	SN	1
E_T^{miss}	SN	3
***b*-tagging**		
Efficiency	SN	30
Mis-tag rate (c)	SN	15
Mis-tag rate (light)	SN	80
Mis-tag rate (extrapolation $c \to \tau$)	SN	1
Signal and background modelling		
Signal		
$t\bar{t}H$ cross-section	N	2
H branching fractions	N	3
$t\bar{t}H$ modelling	SN	1
$t\bar{t}$ Background		
$t\bar{t}$ cross-section	N	1
$t\bar{t}+ \geq 1c$ normalisation	N (free-floating)	1
$t\bar{t}+ \geq 1b$ normalisation	N (free-floating)	1
$t\bar{t}$+light modelling	SN	3
$t\bar{t}+ \geq 1c$ modelling	SN	4
$t\bar{t}+ \geq 1b$ modelling	SN	13
Other Backgrounds		
$t\bar{t}W$ cross-section	N	2
$t\bar{t}Z$ cross-section	N	2
$t\bar{t}W$ modelling	SN	1
$t\bar{t}Z$ modelling	SN	1
Single top cross-section	N	3
Single top modelling	SN	5
W+jets normalisation	N	3
Z+jets normalisation	N	3
Diboson normalisation	N	1
Fakes and non-prompt normalisation	N	7
$t\bar{t}t\bar{t}$ cross-section	N	1
Small background cross-sections	N	9

Table 8.4: The sources of systematic uncertainty on the $t\bar{t}$+jets background in the $t\bar{t}H, H \to b\bar{b}$ analysis.

Systematic source	Description	$t\bar{t}$ categories
$t\bar{t}$ cross-section	Up or down by 6%	All, correlated
$k(t\bar{t}+\geq 1c)$	Free-floating $t\bar{t}+\geq 1c$ normalisation	$t\bar{t}+\geq 1c$
$k(t\bar{t}+\geq 1b)$	Free-floating $t\bar{t}+\geq 1b$ normalisation	$t\bar{t}+\geq 1b$
SHERPA5F vs. nominal	Related to the choice of NLO event generator	All, uncorrelated
PS & hadronisation	POWHEG+HERWIG7 vs. POWHEG+PYTHIA8	All, uncorrelated
ISR / FSR	Variations of μ_R, μ_F, h_{damp} and A14 Var3c parameters	All, uncorrelated
$t\bar{t}+\geq 1c$ ME vs. inclusive	aMC@NLO+HERWIG++: ME prediction (3F) vs. incl. (5F)	$t\bar{t}+\geq 1c$
$t\bar{t}+\geq 1b$ SHERPAOPENLOOPS vs. nominal	Comparison of $t\bar{t}+b\bar{b}$ NLO (4F) vs. POWHEG+PYTHIA8 (5F)	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ renorm. scale	Up or down by a factor of two	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ resumm. scale	Vary μ_Q from $H_T/2$ to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ global scales	Set μ_Q, μ_R, and μ_F to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ shower recoil scheme	Alternative model scheme	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (MSTW)	MSTW vs. CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ PDF (NNPDF)	NNPDF vs. CT10	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ UE	Alternative set of tuned parameters for the underlying event	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ MPI	Up or down by 50%	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 3b$ normalisation	Up or down by 50%	$t\bar{t}+\geq 1b$

8.5 Results

To establish the inclusion of the boosted channel in the analysis, fits are first performed on the single lepton regions with Asimov data, for four different scenarios: boosted only, resolved-only, resolved only with boosted events vetoed, and for resolved+boosted with boosted events vetoed from resolved. The resulting signal strengths for each of these scenarios are summarised in Table 8.5. With a total of five signal regions and six control regions, the resolved analysis represents most of the constraining power of the analysis. Nonetheless, the boosted channel is shown to provide an improvement in the expected overall result by around 4%, compared to analysing those events using the resolved analysis only. This is evaluated by comparing the expected uncertainty on Asmiov fits both with and without the boosted channel. The boosted channel is thus included in the final data fits.

Table 8.5: Signal strengths measured in Asimov data for the four scenarios of how to treat boosted events. The error on μ is reduced when including the boosted region, compared to the resolved only case.

Scenario	μ
Boosted Only	$1.00^{+2.96}_{-3.30}$
Resolved Only	$1.00^{+0.64}_{-0.61}$
Resolved - Boosted	$1.00^{+0.65}_{-0.62}$
Resolved + Boosted	$1.00^{+0.62}_{-0.59}$

In the full fit to data, including the boosted region, the freely floating parameters k_b and k_c obtain best fit values of $k_b = 1.24 \pm 0.10$ and $k_c = 1.63 \pm 0.23$. The observed best fit value for μ is:

$$\mu = 0.84 \pm 0.29(\text{stat})^{+0.57}_{-0.54}(\text{syst}) = 0.84^{+0.64}_{-0.61}.$$

This corresponds to an excess of events over the SM background with an observed (expected) significance of 1.4 (1.6) standard deviations, and a signal strength of $\mu > 2.0$ is excluded at the 95% confidence level. In the absence of signal, the expected exclusion would be $\mu > 1.2$. The distribution of $\log_{10} S/B$, the logarithm of the signal to background ratio for all selected events, is shown in Figure 8.8, and compared to the predictions for the best fit value of $\mu = 0.84$, the excluded value of $\mu = 2.0$, and the background only hypothesis $\mu = 0$.

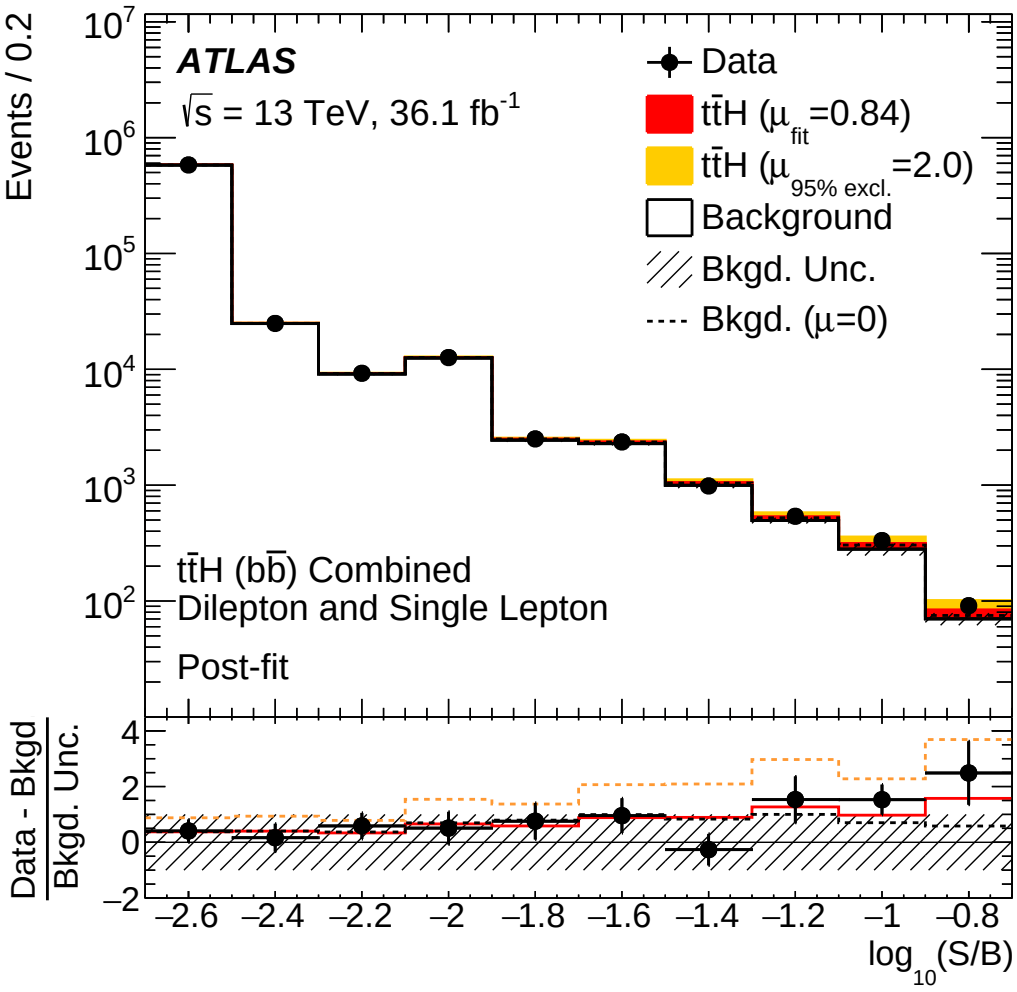


Figure 8.8: The observed number of events in the $t\bar{t}H, H \to b\bar{b}$ analysis as a function of $\log_{10} S/B$, with S the number of signal events and B the number of background events extracted from the profile likelihood fit, compared to the background only hypothesis as well as the observed best fit ($\mu = 0.84$) and excluded value $\mu = 2.0$.

Table 8.6: The post-fit uncertainties in the $t\bar{t}H, H \to b\bar{b}$ analysis. The analysis is dominated by systematics, in particular those due to the $t\bar{t}$+≥1b modelling.

Uncertainty source	$\Delta\mu$	
$t\bar{t}$+≥1b modelling	+0.46	−0.46
Background-model statistical uncertainty	+0.29	−0.31
b-tagging efficiency and mis-tag rates	+0.16	−0.16
Jet energy scale and resolution	+0.14	−0.14
$t\bar{t}H$ modelling	+0.22	−0.05
$t\bar{t}$+≥1c modelling	+0.09	−0.11
JVT, pileup modelling	+0.03	−0.05
Other background modelling	+0.08	−0.08
$t\bar{t}$+light modelling	+0.06	−0.03
Luminosity	+0.03	−0.02
Light lepton (e, μ) id., isolation, trigger	+0.03	−0.04
Total systematic uncertainty	+0.57	−0.54
$t\bar{t}$+≥1b normalisation	+0.09	−0.10
$t\bar{t}$+≥1c normalisation	+0.02	−0.03
Intrinsic statistical uncertainty	+0.21	−0.20
Total statistical uncertainty	+0.29	−0.29
Total uncertainty	+0.64	−0.61

The breakdown of the systematic uncertainties in the analysis are shown in Table 8.6. The largest contribution comes from the modelling of the dominant $t\bar{t}$+≥1b background, and it is this which is the limiting factor on the analysis. Other sub-dominant contributions come from the background model statistical uncertainties, the flavor tagging uncertainties, and the jet energy scale and resolution uncertainties. The statistical uncertainties on the data have a small but non-negligible effect on the final result. The pulls and constraints of the largest 20 systematic uncertainties are shown in Figure 8.9. No pulls beyond 1σ are observed. The largest four systematic uncertainties are all the result of $t\bar{t}$+≥1b modelling, even with large constraints from the fit, particularly on the comparisons to the Sherpa5F and Sherpa4F predictions. Though including both of these comparisons as systematics is likely double counting some systematic effects, this conservative approach ensures that the uncertainty due to the modelling is fully covered. The largest non-modelling systematic is the first nuisance parameter due to light jet mistagging, due to the reliance on b-tagging information in the fit region definitions. This analysis was published in Phys. Rev. D in 2018 [2].

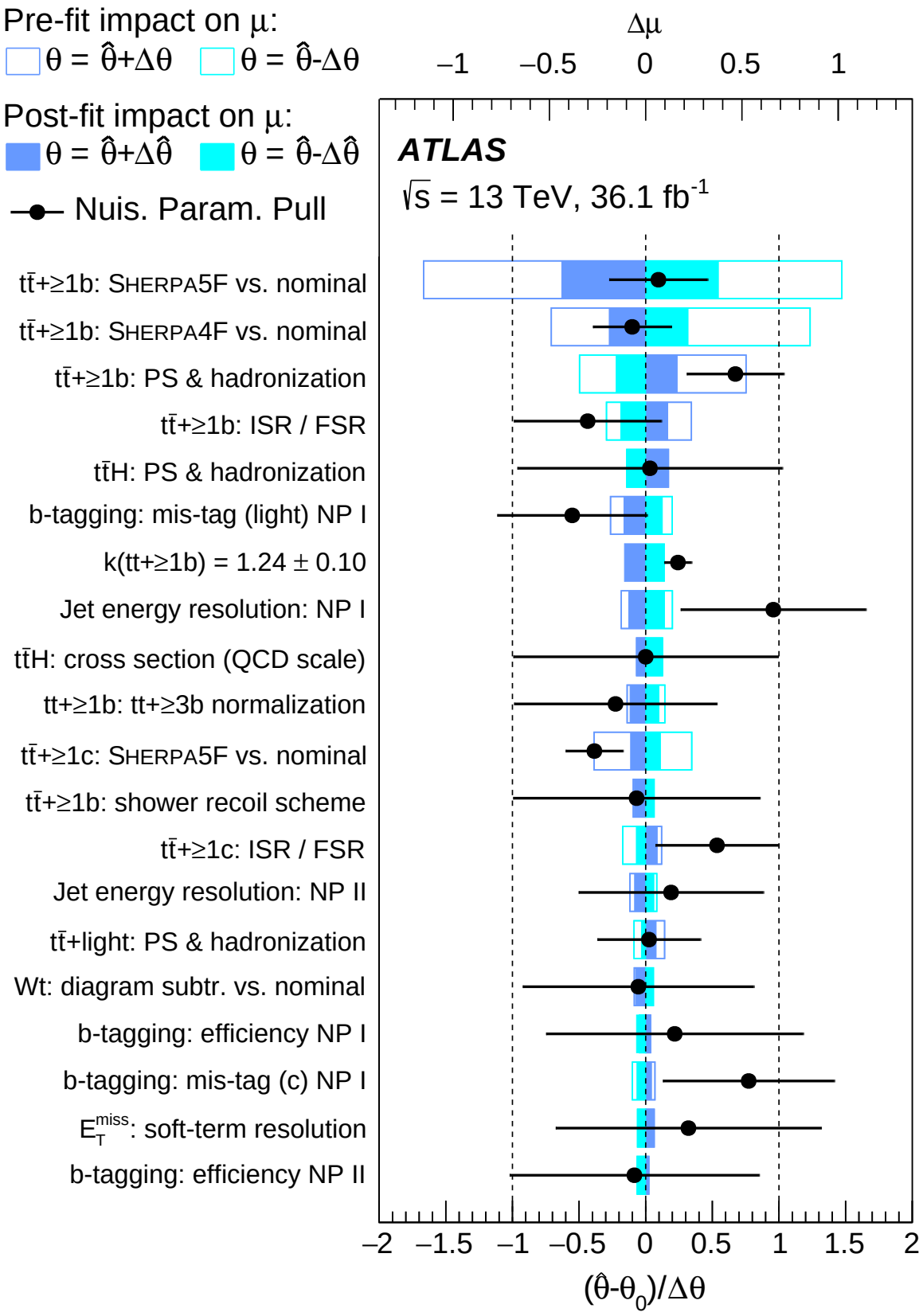


Figure 8.9: The pulls and constraints of the 20 largest systematic uncertainties post-fit. The effect of the nuisance parameters on the final result $\Delta\mu$, calculated by comparing the nominal result with the result after fixing the parameter to it's best fit value $\hat{\theta}$, are shown both pre and post fit by the unfilled and filled rectangles respectively. The pulls of the nuisance parameters from the nominal values θ_0 are shown by the deviation of the black points from 0, with 1σ pull marked by dashed lines.

Table 8.7: Summary of the impact of each group of systematic uncertainties in the Run 2 $t\bar{t}H$ combination.

Uncertainty source	$\Delta\sigma_{t\bar{t}H}/\sigma_{t\bar{t}H}$ [%]
Theory uncertainties (modelling)	11.9
$t\bar{t}$ + heavy flavour	9.9
$t\bar{t}H$	6.0
Non-$t\bar{t}H$ Higgs boson production	1.5
Other background processes	2.2
Experimental uncertainties	9.3
Fake leptons	5.2
Jets, E_T^{miss}	4.9
Electrons, photons	3.2
Luminosity	3.0
τ-leptons	2.5
Flavour tagging	1.8
MC statistical uncertainties	4.4

8.6 Full $t\bar{t}H$ Combination

This $t\bar{t}H, H \to b\bar{b}$ result, including the boosted channel, is subsequently included in a combination [227] with the 36 fb^{-1} $t\bar{t}H$ multilepton analysis which targets Higgs boson decays to $W^{\pm}W^{\mp *}$, $\tau^+\tau^-$, and ZZ^* (excluding $ZZ^* \to 4\ell$) [226], as well as 80 fb^{-1} results for $t\bar{t}H, H \to \gamma\gamma$ and $t\bar{t}H, H \to ZZ^* \to 4\ell$ [227]. The event selections in the analyses are designed to be orthogonal. The combination is performed with another profile likelihood fit, and the correlation of each systematic uncertainty is evaluated separately. In particular, the $t\bar{t}$ modelling uncertainties from the $t\bar{t}H, H \to b\bar{b}$ analysis are considered uncorrelated with the other analyses, as this analysis probes a more extreme region of phase space for this background. Most of the experimental systematics, including most of the jet energy scale, lepton identification/isolation, and trigger efficiencies, as well as signal theory uncertainties, are correlated between all channels. The flavor tagging uncertainties are correlated between the two 36 fb^{-1} analyses and the two 80 fb^{-1} analyses, but uncorrelated between the analyses using different datasets due to updated calibrations being used in the 80 fb^{-1} analyses. The impact of the different systematic uncertainties in the combination is summarised in Table 8.7, where again the $t\bar{t}$+HF uncertainties have the largest impact. The constraints on the flavor tagging uncertainties from the $b\bar{b}$ analysis lead to a reduction in these uncertainties on the multilepton analysis in the combination.

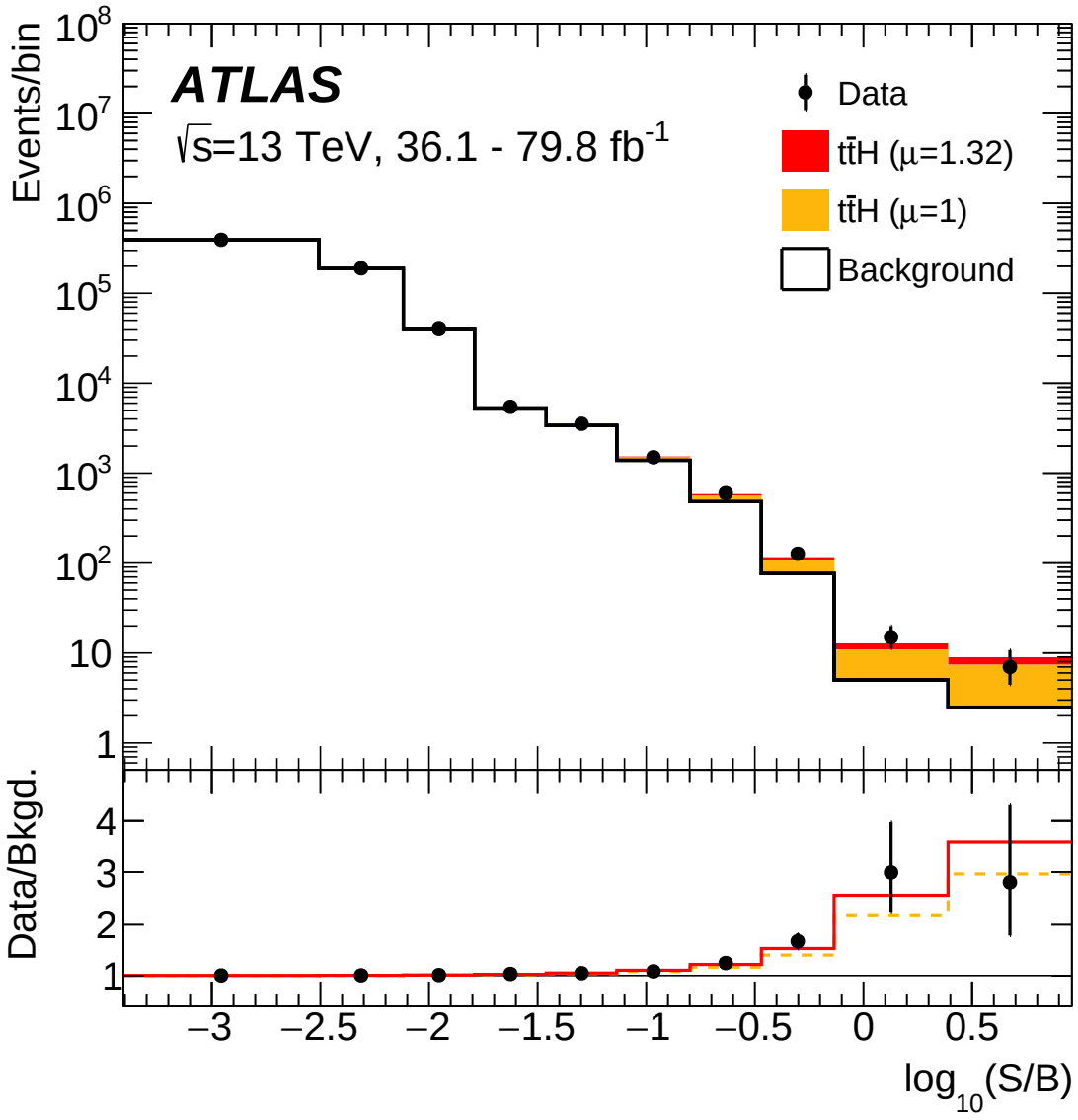


Figure 8.10: The observed number of events in the combined $\sqrt{s} = 13$ TeV $t\bar{t}H$ analyses, as a function of $\log_{10} S/B$, with S the number of signal events and B the number of background events extracted from the profile likelihood fit, compared to the background only hypothesis as well as the SM $t\bar{t}H$ prediction ($\mu = 1$) and the observed best fit ($\mu = 1.32$).

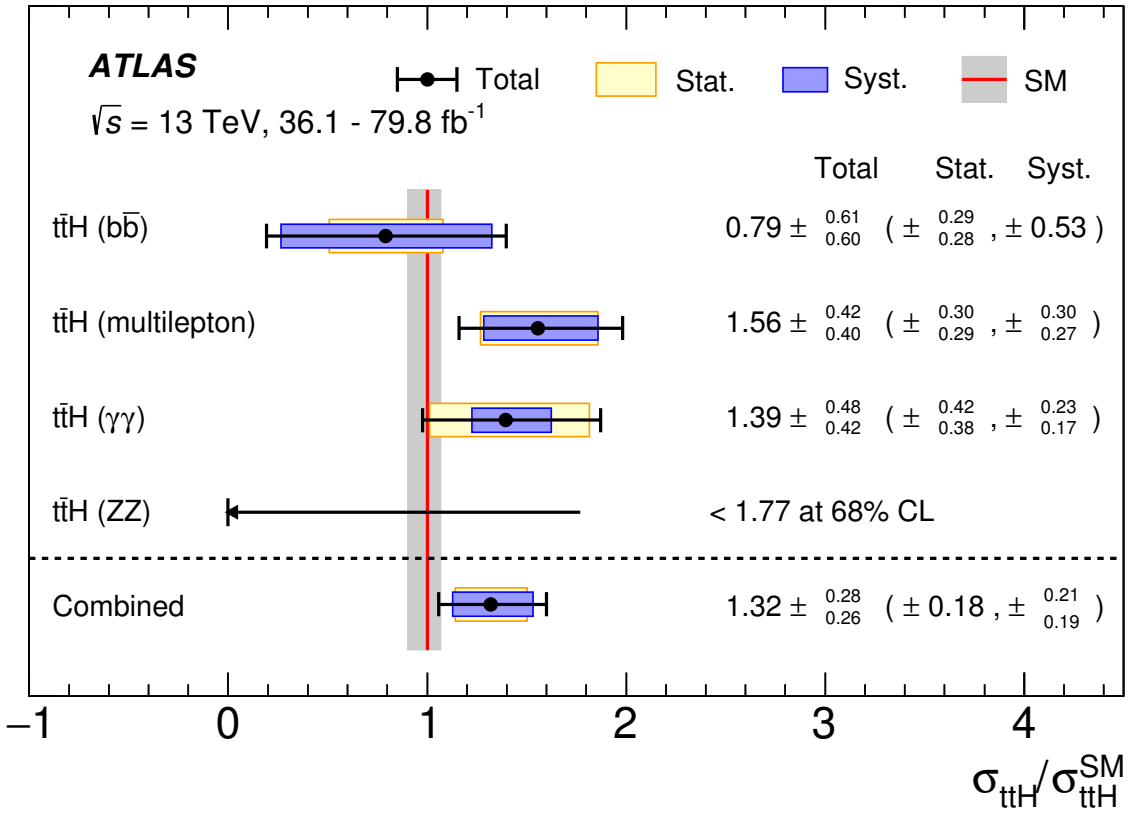


Figure 8.11: The measured signal strengths in each individual analysis, as well as the full $\sqrt{s} = 13$ TeV $t\bar{t}H$ combination. The $t\bar{t}H, H \to b\bar{b}$ and $t\bar{t}H$ multilepton analyses use 36 fb^{-1} of data, while the $t\bar{t}H, H \to \gamma\gamma$ and $t\bar{t}H, H \to ZZ^* \to 4\ell$ analyses are updated to 80 fb^{-1}.

The final result of the combination is an observed (expected) significance of 5.8 (4.9) standard deviations, and thus represents an observation of $t\bar{t}H$ production. The measured cross-section is $670 \pm 90(\text{stat})^{+110}_{-100}(\text{syst})$ fb, which is in agreement with the SM prediction of 507^{+35}_{-50}fb [6,198–202] at NLO in QCD and EW. Figure 8.10 shows the observed event yields in all of the results included in the combination as a function of $\log_{10} S/B$ for the combined 13 TeV results, showing a clear excess of events over the background at high values. The measured value of $\mu = \sigma_{t\bar{t}H}/\sigma^{SM}_{t\bar{t}H}$ in each analysis, as well as the combination, is shown in Figure 8.11. As a result of the correlations, the result shown in Figure 8.11 does not exactly match that quoted in the previous section, with a central value for $t\bar{t}H, H \to b\bar{b}$ of $\mu = 0.79^{+0.61}_{-0.60}$ in the full combination.

Furthermore, a combination with the Run 1 results [41–43] is also performed [227]. The signal modelling uncertainties, as well as the electron/photon energy scales and resolutions, are considered correlated, while the remaining uncertainties are treated as uncorrelated between different centre of mass energies. The total, combined significance for the observation of the $t\bar{t}H$ production process is 6.3 (5.1) observed (expected) standard deviations. The results are summarised in Table 8.8. This is the first direct evidence of the Higgs boson coupling to quarks, an important milestone in the Standard Model description of the origin of quark masses. The top

Table 8.8: Summary of the $\sqrt{s}$ = 13 TeV analyses and combination as well as the Run 1+Run 2 $t\bar{t}H$ combination, which reaches an expected sensitivity of $> 5\sigma$.

Analysis	Integrated luminosity [fb^{-1}]	$t\bar{t}H$ cross section [fb]	Obs. sign.	Exp. sign.
$H \to \gamma\gamma$	79.8	$710\ ^{+210}_{-190}$ (stat.) $^{+120}_{-90}$ (syst.)	4.1σ	3.7σ
$H \to$ multilepton	36.1	790 ± 150 (stat.) $^{+150}_{-140}$ (syst.)	4.1σ	2.8σ
$H \to b\bar{b}$	36.1	$400\ ^{+150}_{-140}$ (stat.) $\pm$ 270 (syst.)	1.4σ	1.6σ
$H \to ZZ^* \to 4\ell$	79.8	< 900 (68% CL)	0.0σ	1.2σ
Combined (13 TeV)	36.1–79.8	670 ± 90 (stat.) $^{+110}_{-100}$ (syst.)	5.8σ	4.9σ
Combined (7, 8, 13 TeV)	4.5, 20.3, 36.1–79.8	–	6.3σ	5.1σ

quark coupling is of particular interest due to its position as the heaviest particle in the SM, and this observation thus represents the achievement of one of the primary objectives of the ATLAS collaboration in Run 2.

The $t\bar{t}H, H \to b\bar{b}$ result was also included in the $H \to b\bar{b}$ combination, along with the VH and ggF production modes, resulting in the first observation of the $H \to b\bar{b}$ decay with a significance of 5.3 (4.8) observed (expected) standard deviations [228]. The Higgs boson couplings to both third generation quarks is thus established with the help of the $t\bar{t}H, H \to b\bar{b}$ analysis.

Chapter 9

Optimising the Boosted $t\bar{t}H, H \to b\bar{b}$ Channel

They were not the same eyes with which he had last looked out at this particular scene, and the brain which interpreted the images the eyes resolved was not the same brain. There had been no surgery involved, just the continual wrenching of experience.

– Douglas Adams

The analysis presented in Chapter 8 is limited by the modelling of the $t\bar{t}$ background. There are thus two possible routes to improving the analysis: improving this modelling, or improving the event selection / separation between signal and background. A study into the background modelling, by increasing statistics in the important regions of phase space via parton shower enhancements, is presented in Section 6.4.1. What follows here are multiple studies investigating alternative event selections and signal-background separation optimisations in the boosted channel, with an investigation into track jet b-tagging presented in Section 9.1, discussion of reclustered jet substructure in Section 9.2, and a novel jet tagging deep neural network study presented in Section 9.3.

9.1 Track Jet b-Tagging

Due to the dense, high p_T topology of the boosted $t\bar{t}H, H \to b\bar{b}$ signal, one possible optimisation of the analysis is the use of track jets, as defined in Section 5.3.3, for the purpose of b-tagging. Typically, track jets are reconstructed with a smaller radius parameter of $R = 0.2$ than the standard calorimeter jets where $R = 0.4$. In the boosted $t\bar{t}H, H \to b\bar{b}$ analysis, the quarks from the top quark and/or Higgs boson decays are produced at high p_T, and thus the angular separation between them is on average much smaller. The narrower cone possible with track jets may thus lead to better event selections and reconstruction.

9.1.1 Track Jet Performance

The performance of track jets are evaluated in a basic pre-selection region, requiring one isolated electron or muon, one top-tagged large jet, and three additional ($\Delta R > 1.0$) small jets outside the top jet. Note that for this study, large jets (as described in Section 5.3.4 and used in Chapter 7) are utilised, rather than reclustered jets (described in Section 5.3.7 and used in Chapter 8), since the use of track jets for b-tagging removes most of the benefits of the reclustered jets in terms of combination with the resolved channels. For $R = 0.4$ calorimeter jets, the small jets must have $p_T > 25$ GeV, while for track jets $p_T > 10$ GeV is required. Two track jet definitions are used in the performance checks, with both R=0.2 and R=0.4 anti-k_T jets investigated. While the R=0.2 jets are of primary interest, the R=0.4 are also included in order to help decouple the differences due to jet radius and jet constituent. The performance of each jet type is evaluated in terms of the *b-tagging efficiency* as well as the *charm rejection* and *light rejection*, evaluated at the calibrated 70% working point. The b-tagging efficiency ϵ_b is defined as

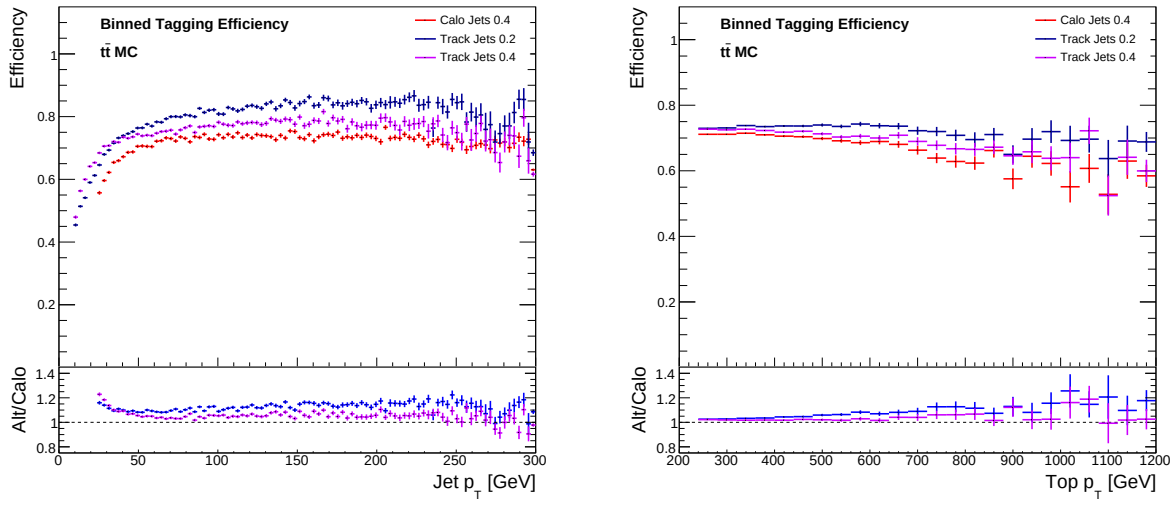
$$\epsilon_b = \frac{N_{b,\text{tag}}}{N_b}, \tag{9.1}$$

where N_b is the number of jets truth matched to a B-hadron (referred to as b-jets) and $N_{b,\text{tag}}$ is the number of those jets that also pass the b-tagging requirement. The light and charm rejections $r_{c,l}$ are defined similarly, as the reciprocal of the tagging efficiency, this time for jets truth matched to either charm or light quarks/gluons, referred to as charm and light jets respectively

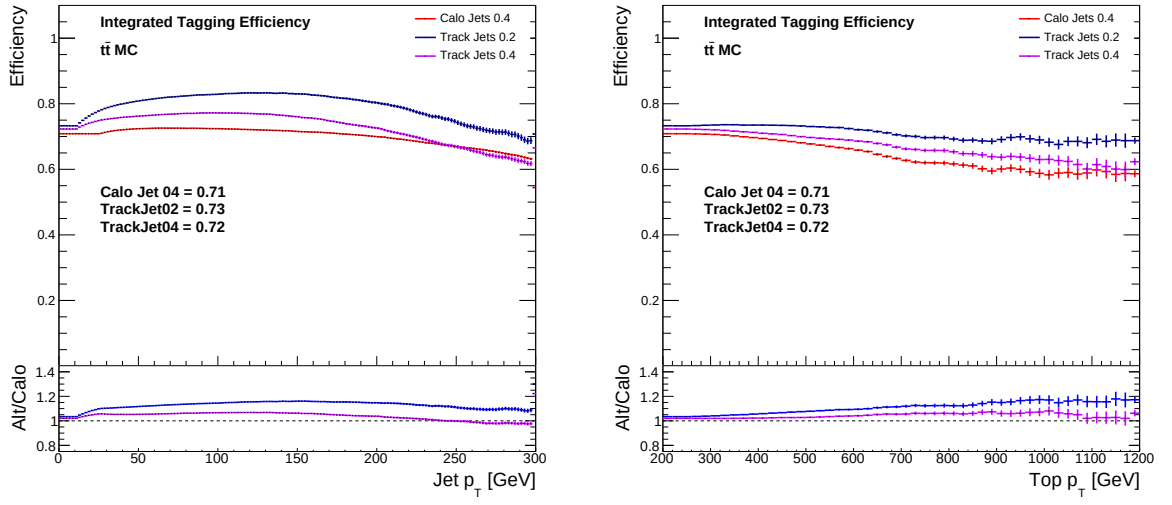
$$r_{c,l} = \frac{1}{\epsilon_{c,l}} = \frac{N_{(c,l)}}{N_{(c,l),\text{tag}}}. \tag{9.2}$$

The tagging efficiencies for b-jets, measured in the nominal $t\bar{t}$ MC sample, are shown in Figure 9.1. It is important to note that the plots showing the tagging efficiency as a function of small jet p_T can be misleading, due to the differing energy scales between the jet collections. For this reason, the tagging efficiency as a function of the top jet p_T is also shown. As expected, the b-tagging efficiency in calorimeter jets decreases as p_T increases more than the track jet efficiency, and in general it is shown that the tagging rate of track jets exceeds that of calorimeter jets in the boosted topoology, despite using the same "70% efficiency" working point. Since the working points are calculated with a dileptonic $t\bar{t}$ selection that is somewhat different to the boosted topology, and the smaller radius jets are designed to perform better in the denser, high energy boosted environment, this is as expected. These results were cross-checked in a $t\bar{t}H$ signal MC sample and found to be compatible.

The tagging efficiencies for b-jets inside the top jet, that is within $\Delta R < 1.0$, are shown in Figure 9.2. From the integrated efficiencies, it can be seen that the $R = 0.2$ jets perform 3% better in the dense environment within the top jet compared to inclusively, while the $R = 0.4$ jets perform similarly in both, showing the advantage of the narrower cone in boosted events.

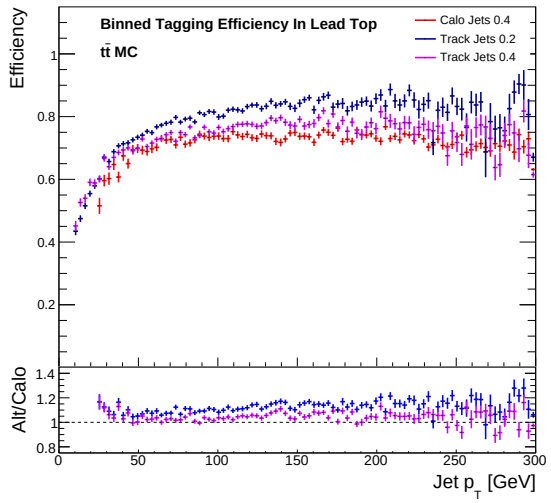


(a) (b)

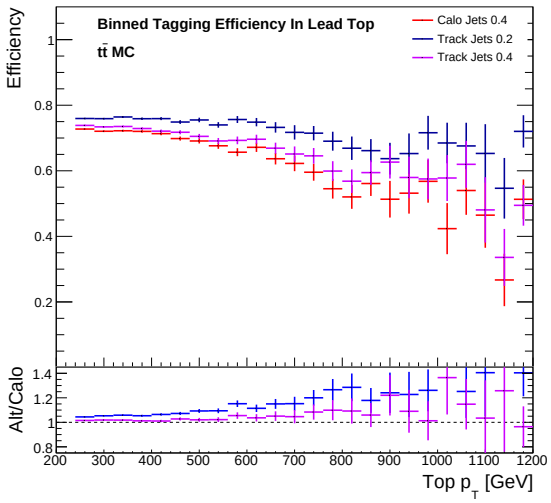


(c) (d)

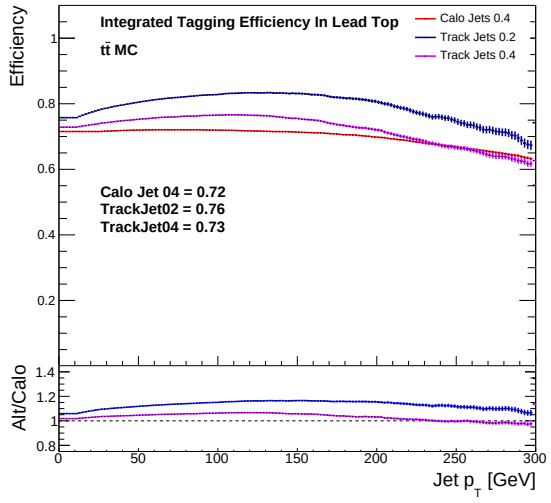
Figure 9.1: Binned (top) and integrated (bottom) tagging efficiencies for b-jets, as a function of small jet p_T (left) and top jet p_T (right), in a $t\bar{t}$ MC sample. For the integrated efficiencies, each bin corresponds to the tagging efficiency for every jet with $p_T > X$ GeV. This also means the initial bins, below the p_T threshold for the small jets, corresponds to the overall efficiency, and this value is shown in the label on these plots. For the binned efficiencies, each bin corresponds to the tagging efficiency for jets only in that range of p_T.



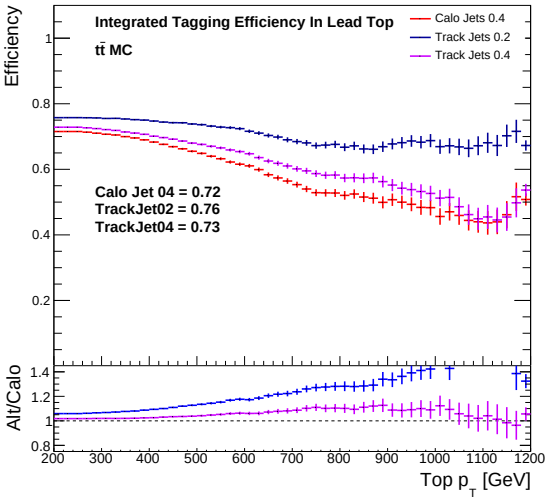
(a)



(b)

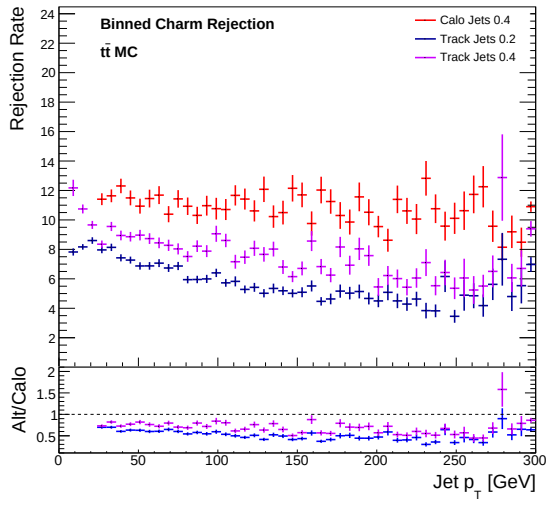


(c)

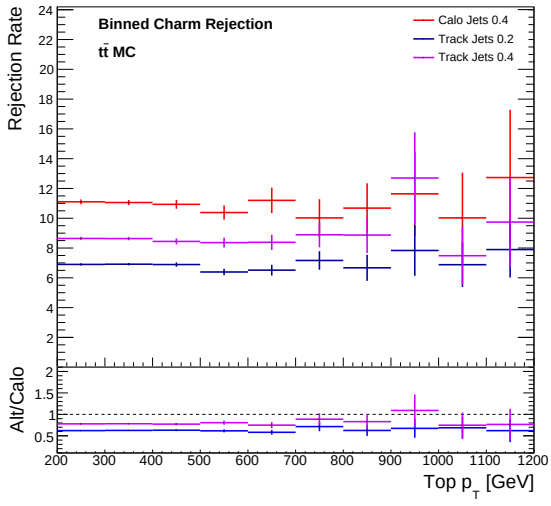


(d)

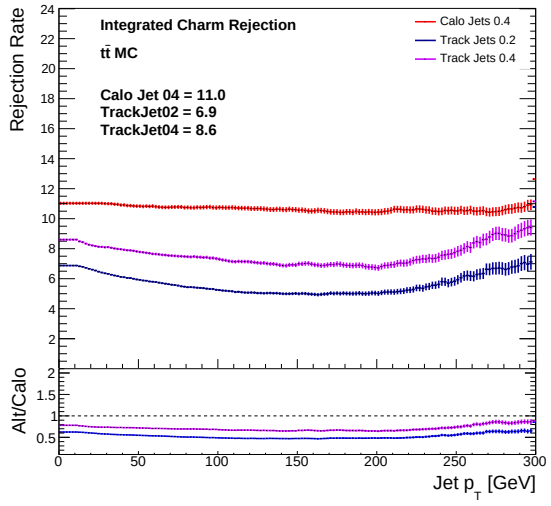
Figure 9.2: Binned (top) and integrated (bottom) tagging efficiencies for b-jets within $\Delta R < 1.0$ of the top jet, as a function of small jet p_T (left) and top jet p_T (right), in a $t\bar{t}$ MC sample. For the integrated efficiencies, each bin corresponds to the tagging efficiency for every jet with $p_T > X$ GeV. This also means the initial bins, below the p_T threshold for the small jets, corresponds to the overall efficiency, and this value is shown in the label on these plots. For the binned efficiencies, each bin corresponds to the tagging efficiency for jets only in that range of p_T.



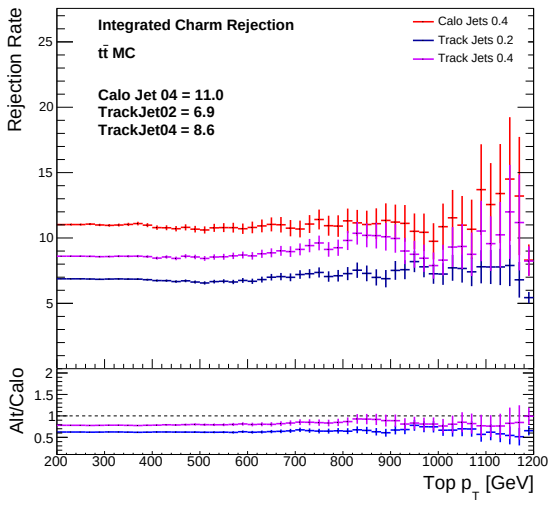
(a)



(b)

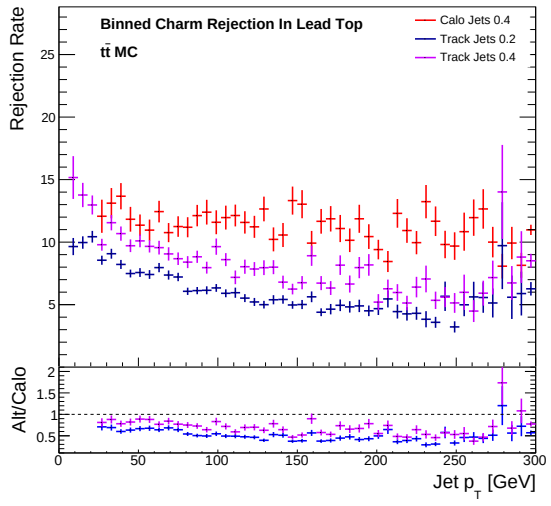


(c)

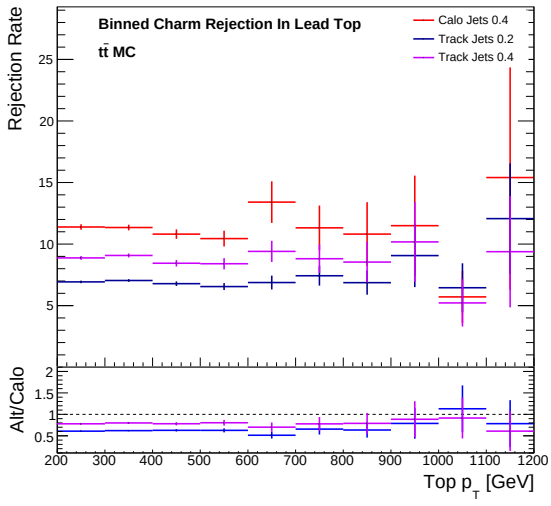


(d)

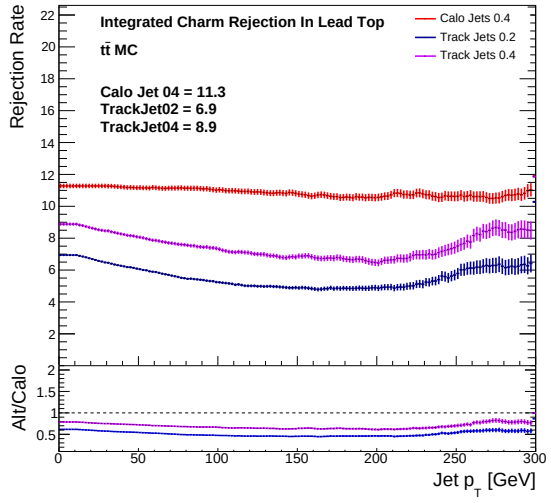
Figure 9.3: Binned (top) and integrated (bottom) rejection rates for charm jets, as a function of small jet p_T (left) and top jet p_T (right), in a $t\bar{t}$ MC sample. For the integrated efficiencies, each bin corresponds to the tagging efficiency for every jet with $p_T > X$ GeV. This also means the initial bins, below the p_T threshold for the small jets, corresponds to the overall efficiency, and this value is shown in the label on these plots. For the binned efficiencies, each bin corresponds to the tagging efficiency for jets only in that range of p_T.



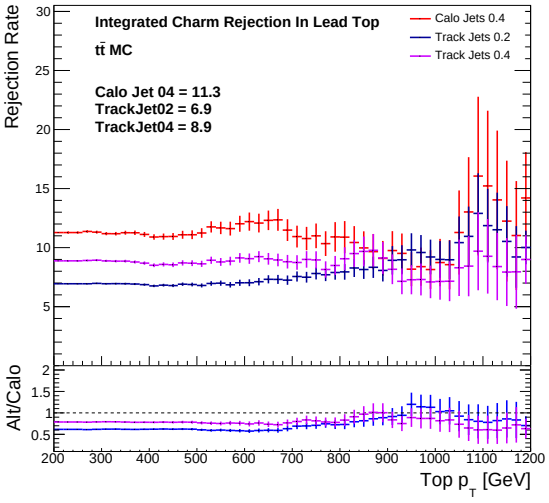
(a)



(b)

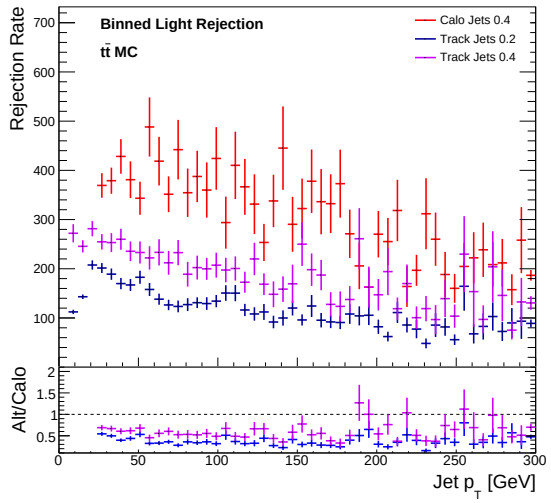


(c)

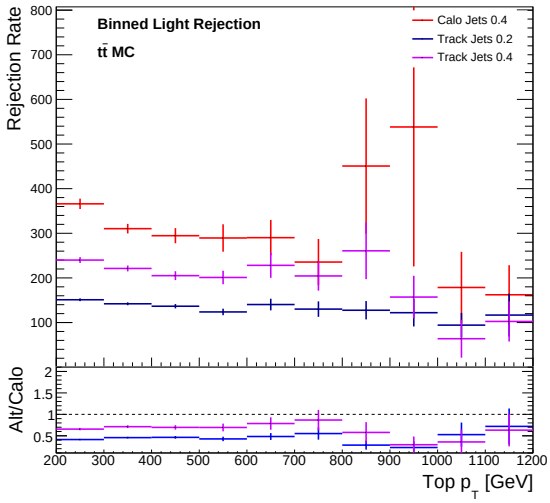


(d)

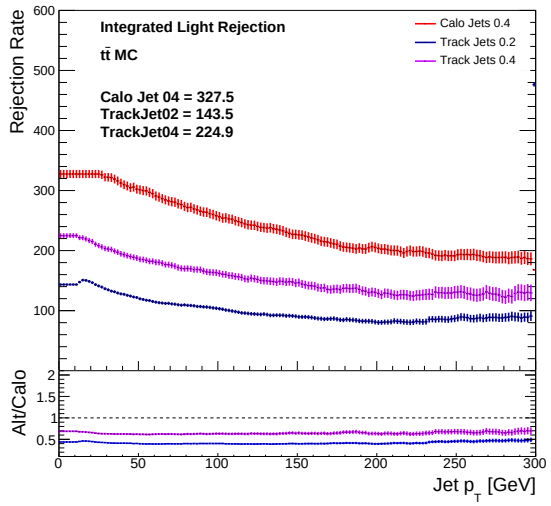
Figure 9.4: Binned (top) and integrated (bottom) rejection rates for charm jets within $\Delta R < 1.0$ of the top jet, as a function of small jet p_T (left) and top jet p_T (right), in a $t\bar{t}$ MC sample. For the integrated efficiencies, each bin corresponds to the tagging efficiency for every jet with $p_T > X$ GeV. This also means the initial bins, below the p_T threshold for the small jets, corresponds to the overall efficiency, and this value is shown in the label on these plots. For the binned efficiencies, each bin corresponds to the tagging efficiency for jets only in that range of p_T.



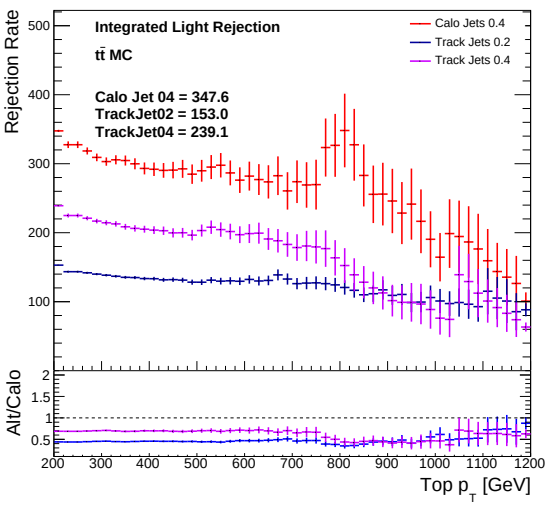
(a)



(b)

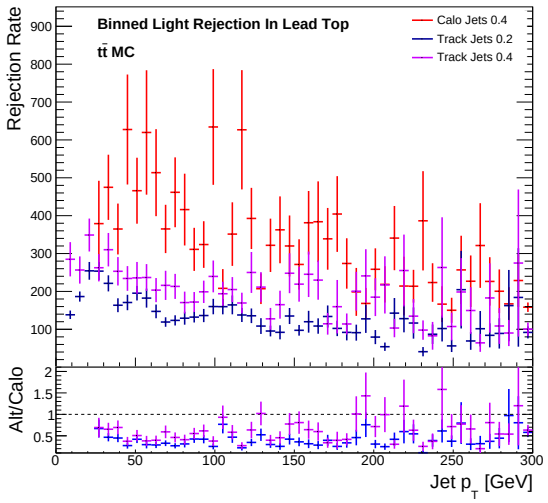


(c)

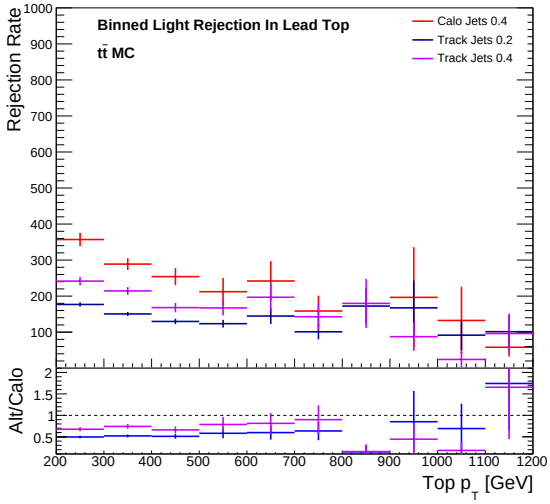


(d)

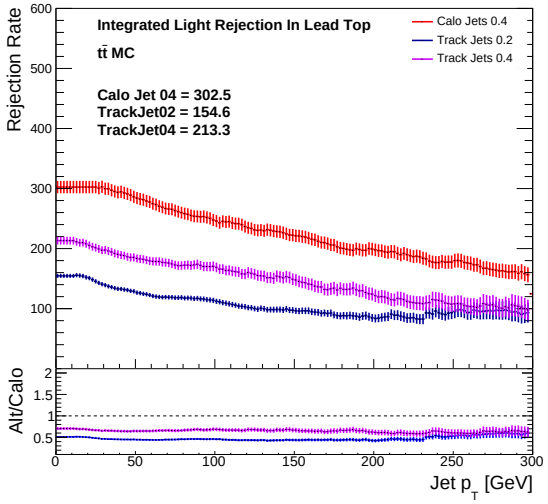
Figure 9.5: Binned (top) and integrated (bottom) rejection rates for light jets, as a function of small jet p_T (left) and top jet p_T (right), in a $t\bar{t}$ MC sample. For the integrated efficiencies, each bin corresponds to the tagging efficiency for every jet with $p_T > X$ GeV. This also means the initial bins, below the p_T threshold for the small jets, corresponds to the overall efficiency, and this value is shown in the label on these plots. For the binned efficiencies, each bin corresponds to the tagging efficiency for jets only in that range of p_T.



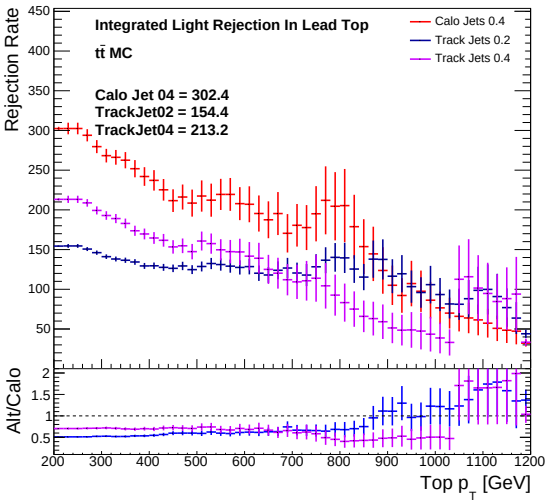
(a)



(b)



(c)



(d)

Figure 9.6: Integrated (top) and binned (bottom) rejection rates for light jets within $\Delta R < 1.0$ of the top jet, as a function of small jet p_T (left) and top jet p_T (right), in a $t\bar{t}$ MC sample. For the integrated efficiencies, each bin corresponds to the tagging efficiency for every jet with $p_T > X$ GeV. This also means the initial bins, below the p_T threshold for the small jets, corresponds to the overall efficiency, and this value is shown in the label on these plots. For the binned efficiencies, each bin corresponds to the tagging efficiency for jets only in that range of p_T.

The charm rejection rates are shown in Figure 9.3 for all charm jets, and in Figure 9.4 for charm jets matched to the top jet. From these it is clear that the rejection is lower in the track jets, particularly in the $R = 0.2$ jets, by around 40%. Similar plots, for the light jet rejection, are found in Figures 9.5 and 9.6, with R=0.2 track jets performing ~50% worse than calorimeter jets here. Since the integrated tagging efficiency was not exactly equal for each jet collection, the rejection rate plots are not quite fair comparisons, since the working point is in effect slightly different for each. By changing the threshold value of the MV2C10 discriminant for each jet collection such that the integrated tagging efficiency is exactly 70%, it was found that the relative mistagging rates remain similar between each jet collection. The implication of this is that using a track jet based selection will result in an increased number of mistagged jets, leading to an increase in the number of background events that will be selected. However, since the $t\bar{t}H$ signal is small compared to the background, maximising the selection efficiency at the cost of increased backgrounds may still be worthwhile if the signal vs background separation is good. To investigate this, two new selections are investigated in Sections 9.1.2 and 9.1.3 to compare track jet and calorimeter jet b-tagging in the $t\bar{t}H, H \to b\bar{b}$ boosted channel, using MC samples corresponding to 36 fb^{-1} of $\sqrt{s} = 13$ TeV data.

9.1.2 High Purity Top-tag Region

The first of the two regions investigated, labelled the *high-purity top-tag* region, targets high p_T top quarks and Higgs bosons of any p_T. This region has the same requirements as the preselection region used in Section 9.1.1, tightened further by requiring that the top-tagged large jet has exactly one b-tagged small jet and at least one non-b-tagged matched small jet. Furthermore, there must be at least three additional b-tagged small jets that are separated from the top jet by $\Delta R > 1.0$. This selection is summarised in Figure 9.7.

In studies of this region using calorimeter jets, a classification BDT was already trained and optimised [229], and this was reused in the study utilising track jets. However, some of the variables used in this training involve an invariant mass, for example m_{bb} or similar, which have degraded signal vs background separation in track jets – track jets can reconstruct energy deposits only from charged particles, and are thus missing the neutral particle contribution to the four vectors. Since these variables are known to be important for signal and background separation, a matching of track jets to calorimeter jets is performed in order to build these variables, instead of developing a dedicated BDT for the track jet selection from scratch which replaces or leaves out these variables. This matching is performed only for b-tagged jets outside the top jet using optimised angular matching such that each track jet is matched to the closest small radius calorimeter jet (with $\Delta R \leq 0.4$), without allowing the same jet to be matched to more than one track jet, using the Hungarian Algorithm [230]. This introduces some amount of inefficiency, both in finding a solution and in rejecting at $\Delta R(\text{track jet}, \text{calorimeter jet}) < 0.4$,

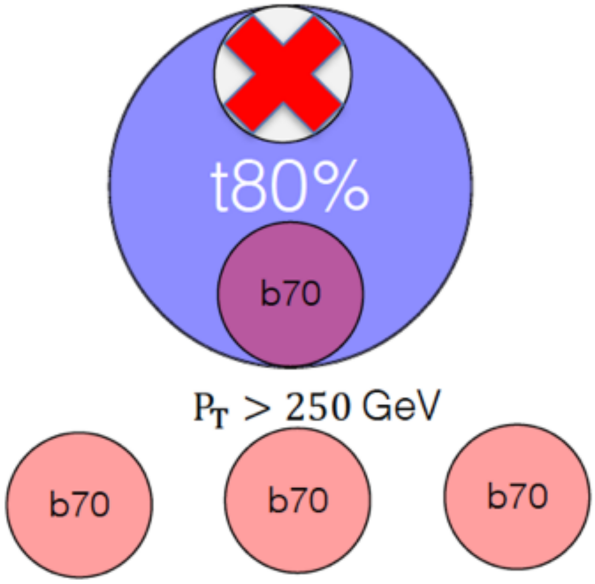


Figure 9.7: Cartoon summary of the high purity top-tag region. The large blue circle represents a top-tagged (at the 80% working point) large jet, while the smaller pink circles labelled "b70" represent the small jets or track jets b-tagged at the 70% WP. The small white circle with the red cross represents a small jet or track jet that does not pass the b-tagging requirement. Small jets inside the large jet are within $\Delta R < 1.0$, while those outside have $\Delta R > 1.0$.

Table 9.1: Yields in the high purity top-tag region, comparing the different jet collections used for b-tagging. The third, fifth and seventh columns show the difference with respect to the calorimeter jet selection.

Selection	Sig.Yield	ΔS (%)	Bkg.Yield	ΔB (%)	$S/\sqrt{B}$	$\Delta(S/\sqrt{B})$ (%)
Calorimeter Jets	14.85	–	197	–	1.06	–
All Track Jets	17.68	+19.07	326	+64.97	0.98	-7.30
Matched Track Jets	13.07	-11.97	174	-12.04	0.99	-6.14

and cases where this matching fails for a particular track jet corresponds to removing that b-tag from the event. This causes some events to be lost, though this effect is present in both signal and background. Once this matching is performed, the relevant variables are in good agreement between the different jet collections in both signal and background and thus there is no need to retrain the MVA.

Table 9.1 shows the signal and background yields, estimated using the same methods as described in Chapter 8, for the calorimeter jet collection and the R=0.2 track jet collection both before and after applying the matching, showing that the signal yield would increase without the matching inefficiency. The overall $S/\sqrt{B}$ does not change much with the matching procedure, since the inefficiency of the matching also acts to suppress the background. Profile likelihood fits are then performed using the same background estimation as described in Chapter 8. Using only

the high purity top-tag region, an Asimov fit is performed to the BDT trained on the calorimeter jet selection, for both the calorimeter jet selection and the matched track jet selection, and a statistics-only limit on μ is obtained. The fit is not performed for the unmatched selection, since many variables are not well defined here. The expected upper limit on $\mu = \frac{\sigma_{t\bar{t}H}}{\sigma_{t\bar{t}H}^{\mathrm{SM}}}$ in the calorimeter jet fit is $\mu > 1.94$, while in the track jet selection it is $\mu > 2.63$. Thus, the track jet selection degrades the limit in this region by approximately 35%. This is consistent with the reduced $S/\sqrt{B}$ caused by the higher mistagging rate, the loss in signal yield due to the matching procedure used, and the lack of dedicated BDT training. While further optimisation, such as via a dedicated optimisation of the MVA or improved matching procedure, may improve the situation, the expected loss is large and so further studies were not performed for this region.

9.1.3 Trimmed Top + Higgs Region

The second region considered for the track jet study targets both a high p_T top quark and high p_T Higgs boson. The same preselection from 9.1.1 is used, with further requirements of a second large jet with $p_T > 200$ GeV, and two b-tagged small jets ΔR matched to this second large jet (called the Higgs-jet). Furthermore, there must be at least one b-jet ΔR matched to the top jet. All of the b-tags in this region use the 85% working point. This selection is summarised in Figure 9.8. This region is similar to the boosted region used in Chapter 8, without the additional b-tag requirement and using large jets in place of reclustered jets, and is referred to as the *trimmed top + Higgs* region to reflect this.

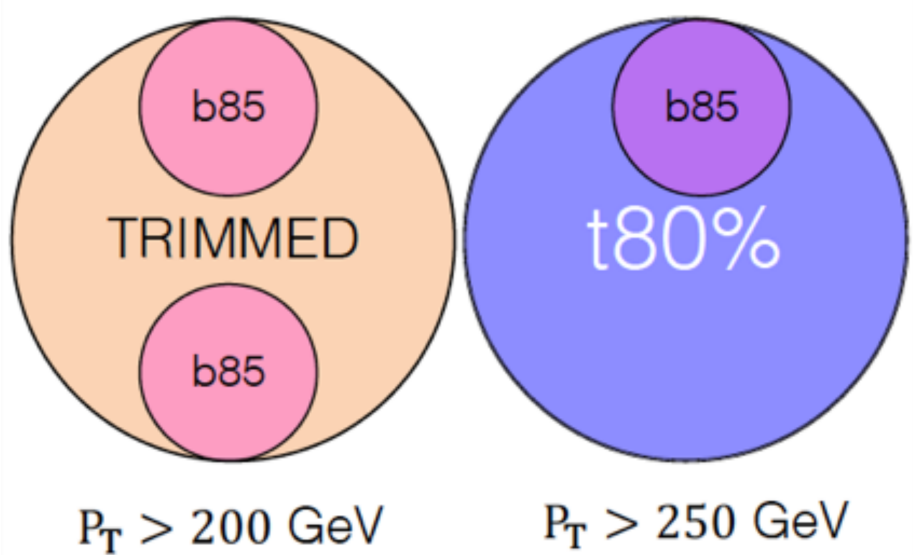


Figure 9.8: Cartoon summary of the trimmed top+Higgs region. The large orange circle represents a large jet, while the blue circle represents a top-tagged (at the 80% working point) large jet. The smaller pink circles labelled "b85" represent the small jets or track jets b-tagged at the 85% WP. Small jets inside either of the large jet are within $\Delta R < 1.0$ of that large jet.

Table 9.2: Yields in the trimmed top+Higgs region. Although the signal yield increases, the $S/\sqrt{B}$ goes down due to the large increase in background, which can be attributed to the increased mistagging rates.

Selection	Sig.Yield	ΔS (%)	Bkg.Yield	ΔB (%)	S/$\sqrt{B}$	Δ($S/\sqrt{B}$) (%)
Calo. Jets	26.8	–	1026.9	–	0.84	–
Track Jets	33.9	+26.5	1772.7	+72.6	0.80	-3.8

The track-jet tagging is particularly appealing in this region for the purpose of the "Higgs-tag", since the narrower cones mean the two b-jets are less likely to overlap and be identified as a single jet. The signal and backgrounds are estimated using the same methods as Chapter 8, and the final yields obtained in this region are shown in Table 9.2. An increased signal yield is accompanied by an even larger increase in the background yield, leading to a degraded $S/\sqrt{B}$, due to the increased mistagging rates. The classification BDT trained in this region [229] contains no variables that require the previously described matching procedure, since variables such as m_{bb} can be replaced with the mass of the Higgs-jet. However, a large amount of tagging information, in the form of b-tagging variables related to the MV2 discriminant, is included in the MVA. Since the discriminants for track jets and calorimeter jets are independent variables, the signal region MVA is retrained using track jet inputs.

A statistics-only Asimov fit is performed to the calorimeter jet and track jet selections, as in Section 9.1.2. The upper limit on μ in the calorimeter jet selection is $\mu > 4.9$ while in the track jet selection it is $\mu > 5.4$. Thus, the overall limit obtained is degraded by ~10%. This can be understood as due to the lower $S/\sqrt{B}$, the unoptimised (though retrained) BDT, and the changed background composition.

9.1.4 Track Jets Conclusions

As demonstrated in Sections 9.1.2 and 9.1.3, the use of track jets for b-tagging was not able to improve the $t\bar{t}H, H \to b\bar{b}$ boosted analysis. However, the findings of this study did lead to improvements in the MVA training and calibration of track jet collections in ATLAS:

- The radius of the cone around the jet used to associate tracks for b-tagging is p_T dependent, and due to the results in this study it was discovered that the same p_T scale was being assumed for calorimeter and track jets. Thus, the radius used in track jets was not optimal, due to the missing neutral particle contributions. This has since been changed by adding a multiplicative factor of 1.6 to approximately correct for the missing contributions.

- The MVA training for the b-tagging algorithms includes the jet p_T as an input variable, but dedicated training was not performed for the track jets. This is now performed in order to correct this slight mismatch.

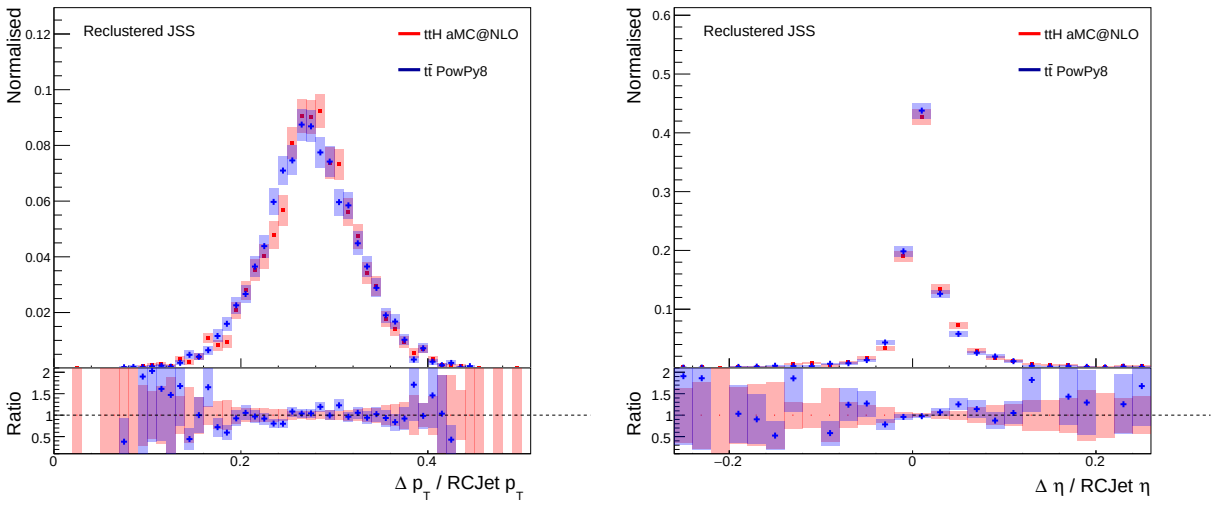
These changes have lead to small improvements in track jet b-tagging performance in ATLAS for subsequent analyses. However, to be useful in $t\bar{t}H, H \to b\bar{b}$, the rejection rates of light and charm need to be improved substantially. For this reason, track jets are not used in the studies utilising the full dataset in the following sections.

9.2 Reclustered Jet Substructure

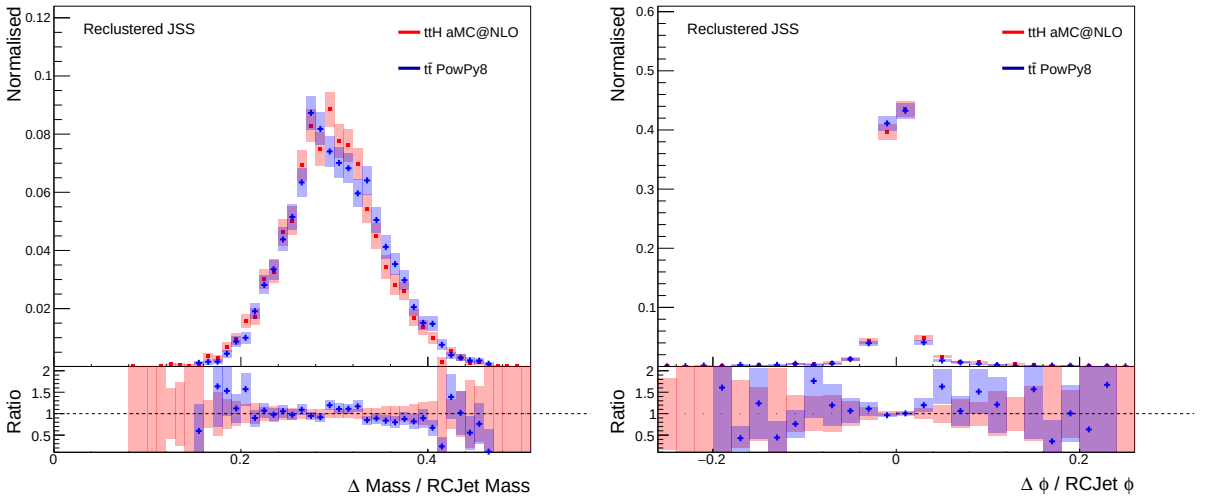
In Section 5.3.7, it was stated that some of the jet substructure variables, defined in Section 3.1, are not well defined for reclustered jets. This is because of the lack of calorimeter cluster information when using small jets as inputs. However, by taking the clusters used to build the small jets, and directly building an R=1.0 anti-k_T "proxy" jet from them, it is possible to calculate variables such as τ_{21}, τ_{32}, and D_2, which can then be used with the reclustered jet. The variables $\sqrt{d_{12}}$, $\sqrt{d_{23}}$ and Q_W, although already defined using subjet inputs, can also be defined in this alternative way. A detailed description of these variables can be found in Section 5.3.5.

9.2.1 Proxy Jet

The proxy jet is not the same as the reclustered jet, as the calibrations which are applied at the small jet level are lost: the differences in the 4-vector between the reclustered jet and the proxy jet are shown in Figure 9.9, for both $t\bar{t}$ and $t\bar{t}H$ MC samples. It can be seen that the jet axis is very close to correct in most cases, as shown by the η and ϕ distributions, but that the energy scale is consistently off by around 30% in the p_T and mass distributions. The clusters used as inputs to small jets are calibrated at the EM scale, rather than the LCW scale usually preferred for large jets (see Section 5.1.1), which causes this shift. Nonetheless, many of the JSS variables are ratios, and so the scale cancels out. The definition also provides pile up robustness, due to the removal of pileup jets via the use of the jet vertex tagger, described in Section 5.3.1, before the reclustering is performed.



(a) (b)



(c) (d)

Figure 9.9: The difference between the reclustered jet and the proxy jet for each component of the jet 4-vector, normalised by the reclustered jet 4-vector. Consistent results are seen in $t\bar{t}H$ MC, shown in red, and $t\bar{t}$ MC, shown in blue.

9.2.2 Systematic Uncertainties

In order to use these variables in a real analysis, systematic uncertainties must be defined for them. To do this, the R_{track} method is used [157], as described in Section 5.3.4, using 80 fb^{-1} of $\sqrt{s} = 13$ TeV data collected between 2015 and 2017.

Events are required to pass large jet triggers, with at least one R=1.0 anti-k_T jet with $p_T >$ 360, 420, or 440 GeV in 2015, 2016 and 2017 data respectively. Lower p_T bins are further populated in 2016 and 2017 using prescaled triggers with $p_T > 260$ GeV. Dijet MC, described in Section 6.8, is used to compare to data. The distributions of r_{track}, the ratio between the substructure variables calculated with calorimeter clusters and ID tracks, and R_{track}, the double ratio of r_{track} in data and MC, are shown per variable in Figure 9.10. These show that the nominal Pythia sample has generally good agreement between data and MC, while the alternate Herwig7 and Sherpa samples have large disagreement with the data in some variables.

The final uncertainties for each of the six variables, which are binned in regions of m/p_T, are shown in Figures 9.11–9.16 for 4 representative bins each. The "baseline" component represents the comparison between data and Pythia MC, while the "modelling" component is the envelope between the nominal and the alternate samples. Furthermore, systematic components are considered on the tracking efficiency, fake rate, and bias, as well as the total statistics in the available samples. In almost all regions of phase space, the modelling component is dominant. In general, the systematic uncertainty tends to grow with p_T, in particular for the τ_{21} variable shown in Figure 9.11, which has the largest uncertainty of the variables under consideration, being as high as 30% in some regions. The τ_{32} variable, shown in Figure 9.12, shows a smaller p_T dependence, and typically has an uncertainty of around 5%.

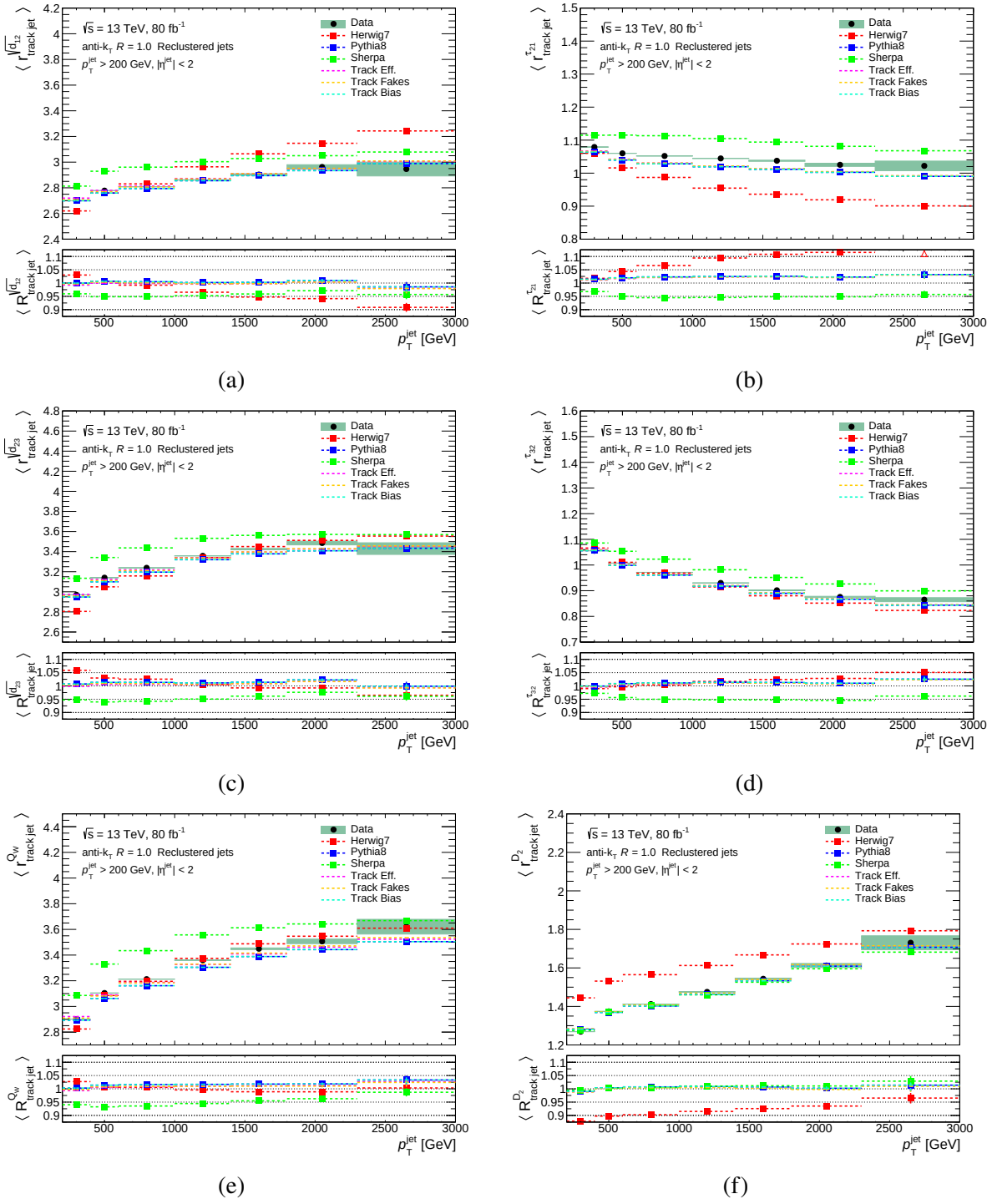


Figure 9.10: Distributions of r_{track} (upper panel) and R_{track} (lower panel) for the six JSS variables, in data (black), the nominal prediction from PYTHIA8 (blue), the alternate predictions from HERWIG7 (red) and SHERPA (green), and the tracking variations in the PYTHIA MC (pink, orange, and cyan).

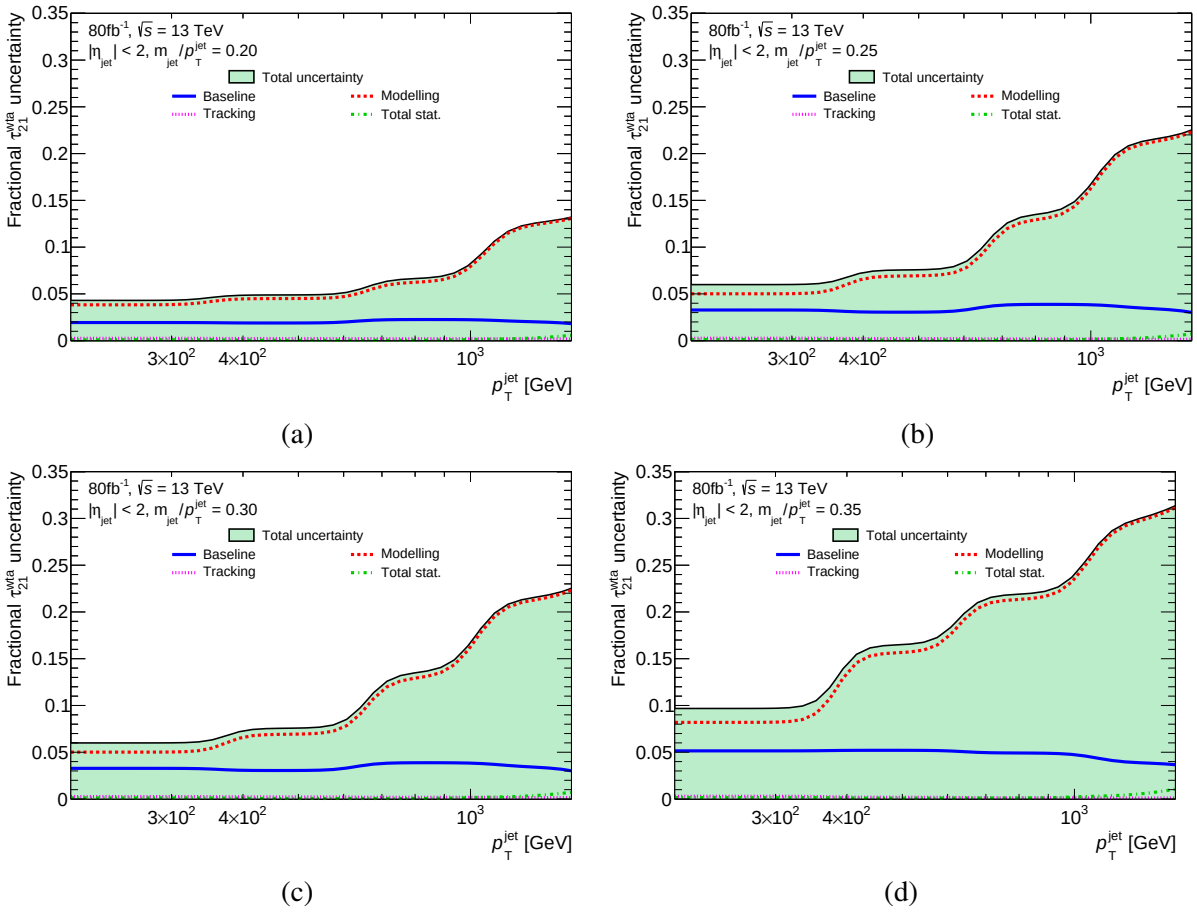


Figure 9.11: The final systematic uncertainties for the τ_{21} variable, in four bins of m/p_T. The total uncertainty is shown in shaded blue, with the individual contributions due to the agreement between data and baseline Pythia MC (solid blue), Pythia and alternate MC (dotted red), tracking variations (dotted pink), and limited statistics (dotted green) also shown.

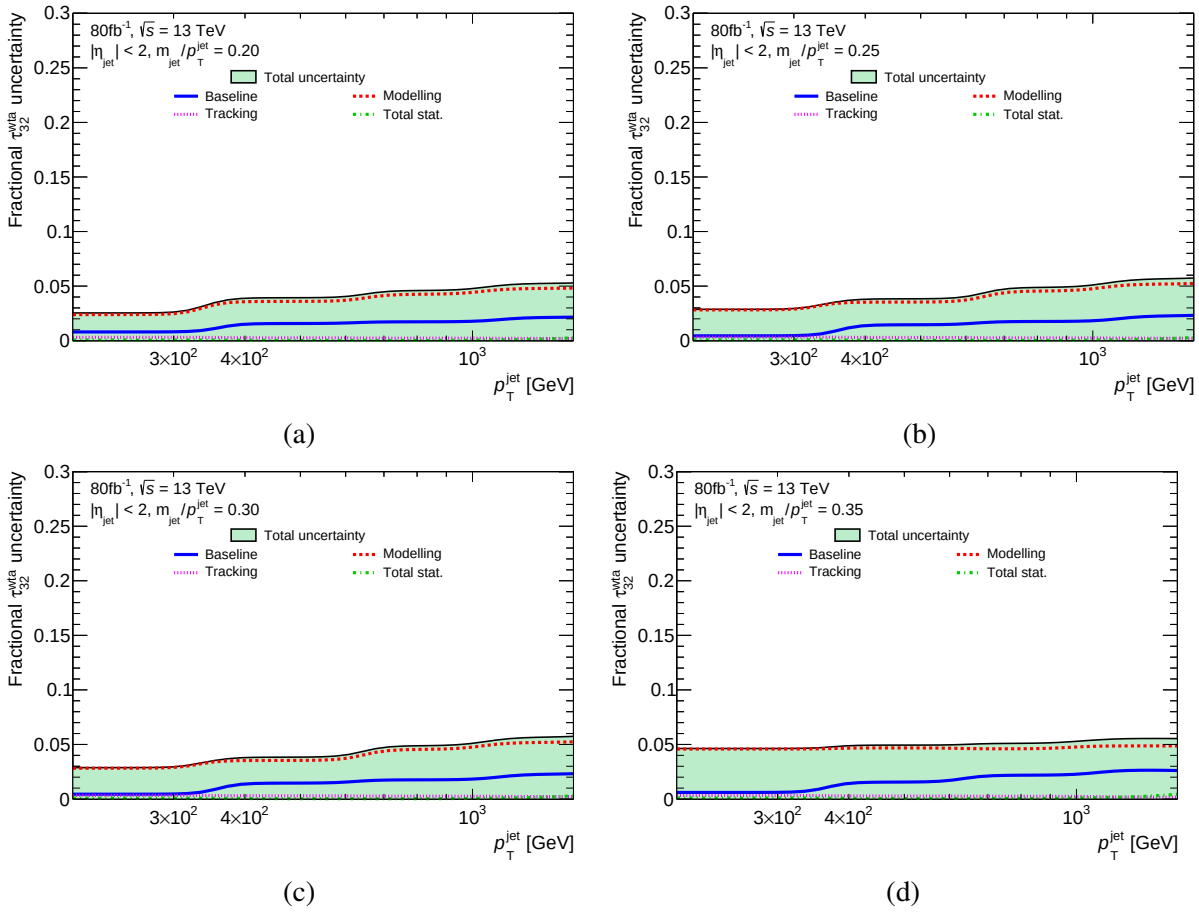


Figure 9.12: The final systematic uncertainties for the τ_{32} variable, in four bins of m/p_T. The total uncertainty is shown in shaded blue, with the individual contributions due to the agreement between data and baseline PYTHIA MC (solid blue), PYTHIA and alternate MC (dotted red), tracking variations (dotted pink), and limited statistics (dotted green) also shown.

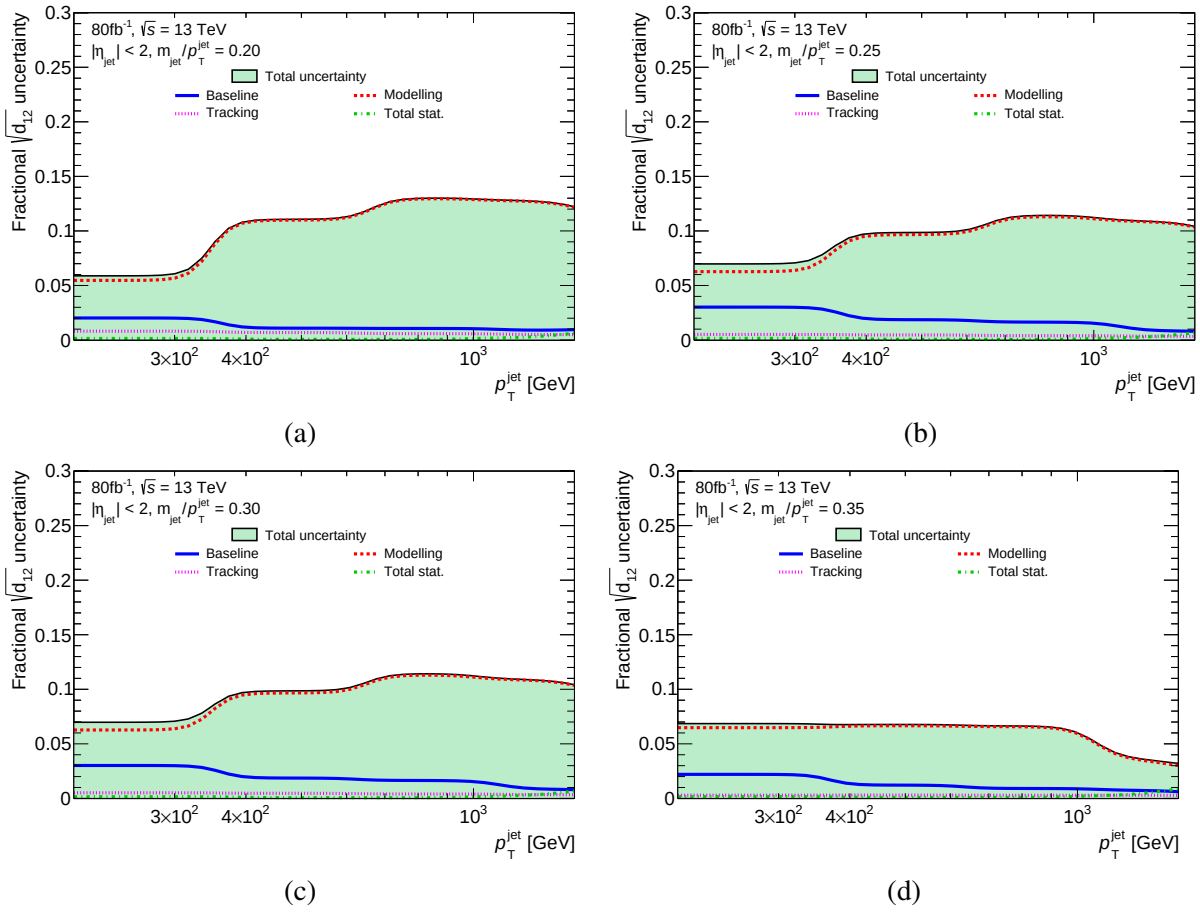


Figure 9.13: The final systematic uncertainties for the $\sqrt{d_{12}}$ variable, in four bins of m/p_T. The total uncertainty is shown in shaded blue, with the individual contributions due to the agreement between data and baseline PYTHIA MC (solid blue), PYTHIA and alternate MC (dotted red), tracking variations (dotted pink), and limited statistics (dotted green) also shown.

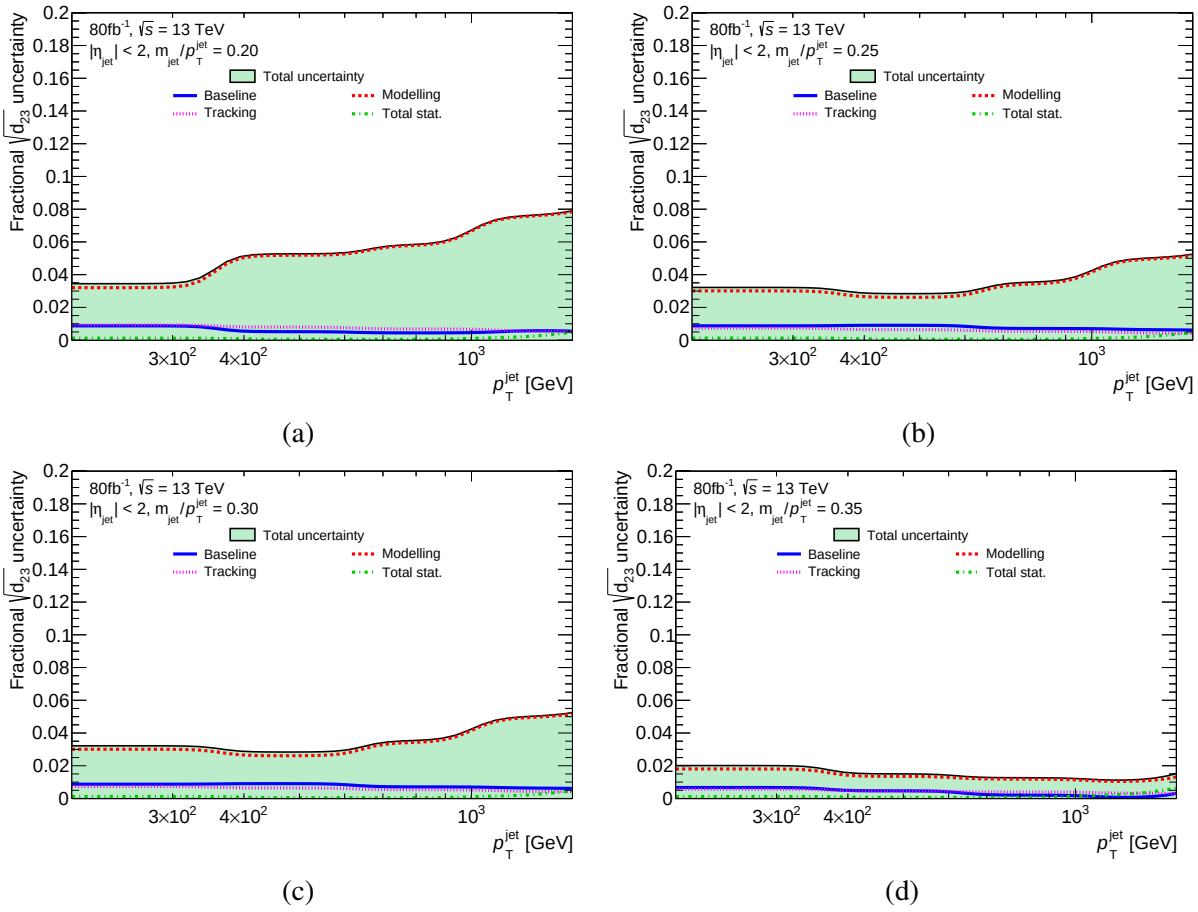


Figure 9.14: The final systematic uncertainties for the $\sqrt{d_{23}}$ variable, in four bins of m/p_T. The total uncertainty is shown in shaded blue, with the individual contributions due to the agreement between data and baseline PYTHIA MC (solid blue), PYTHIA and alternate MC (dotted red), tracking variations (dotted pink), and limited statistics (dotted green) also shown.

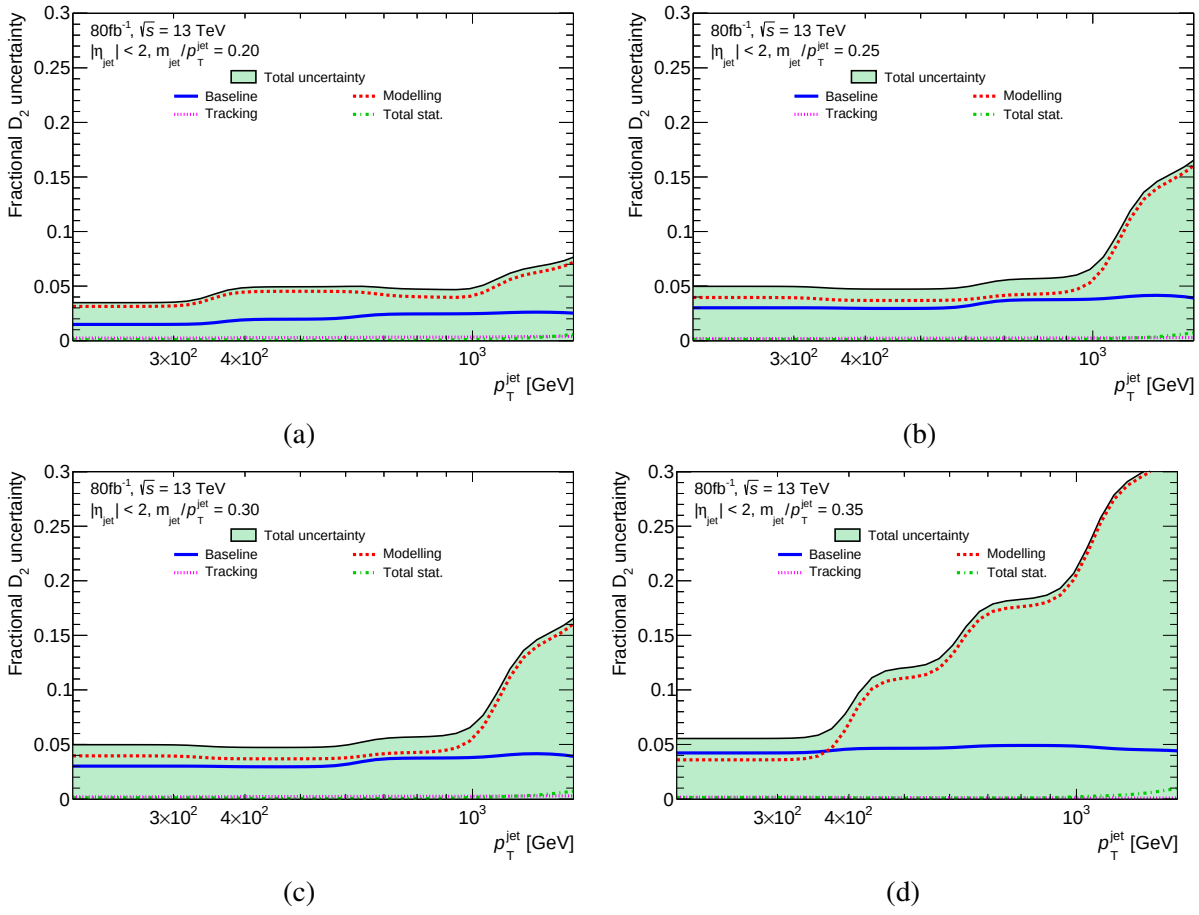


Figure 9.15: The final systematic uncertainties for the D_2 variable, in four bins of m/p_T. The total uncertainty is shown in shaded blue, with the individual contributions due to the agreement between data and baseline PYTHIA MC (solid blue), PYTHIA and alternate MC (dotted red), tracking variations (dotted pink), and limited statistics (dotted green) also shown.

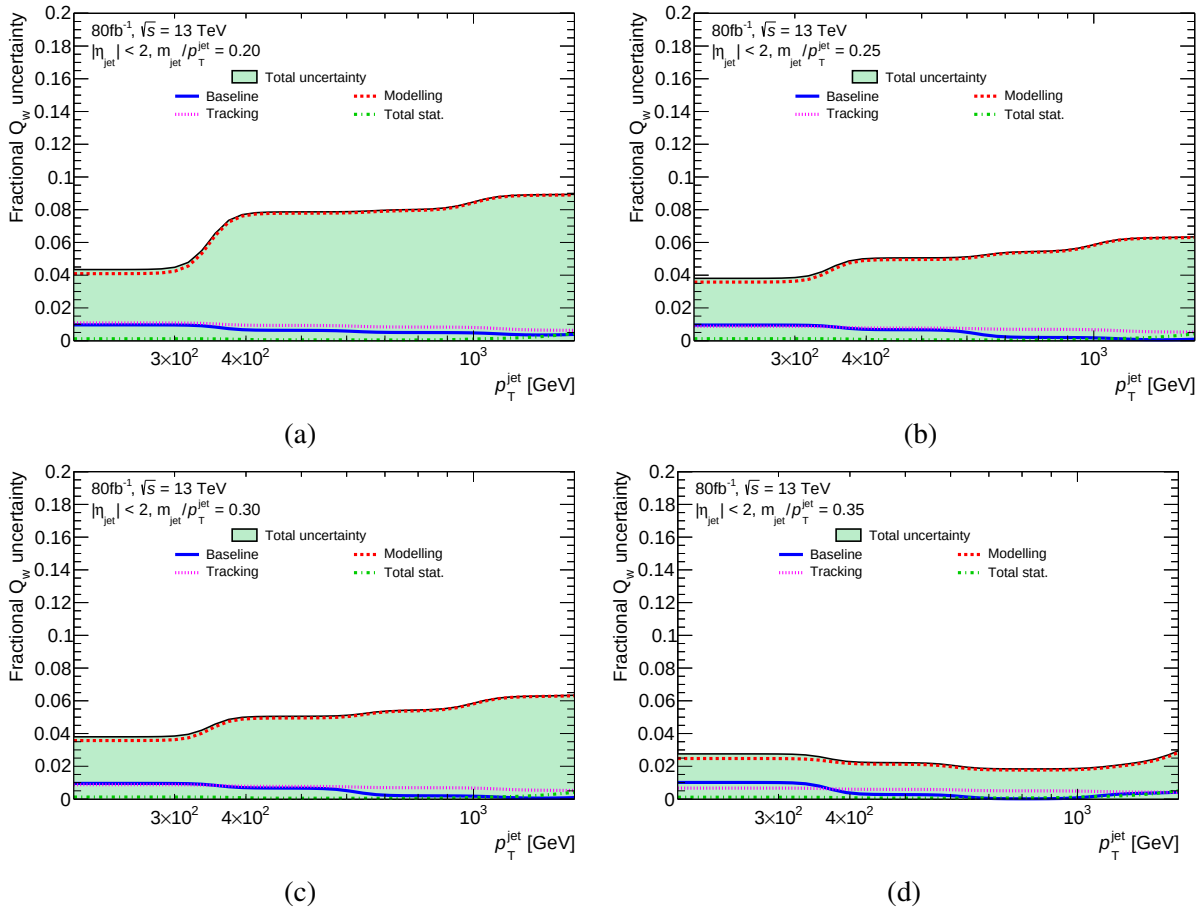


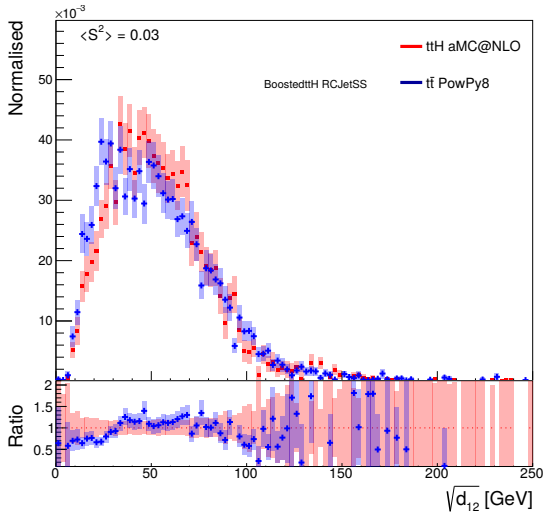
Figure 9.16: The final systematic uncertainties for the Q_W variable, in four bins of m/p_T. The total uncertainty is shown in shaded blue, with the individual contributions due to the agreement between data and baseline Pythia MC (solid blue), Pythia and alternate MC (dotted red), tracking variations (dotted pink), and limited statistics (dotted green) also shown.

9.2.3 Application to $t\bar{t}H, H \to b\bar{b}$

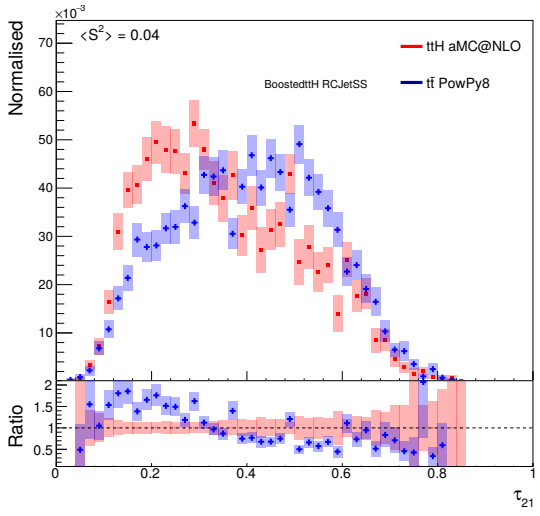
An example application for these variables is in the boosted $t\bar{t}H, H \to b\bar{b}$ analysis, as input variables to the classification BDT. Plots comparing the distribution in $t\bar{t}H$ signal and $t\bar{t}$ background MC samples, for the Higgs jet in the boosted selection from Chapter 8, are shown for each variable in Figure 9.17. These plots demonstrate that the distribution of each of the jet substructure variables is different between signal and background; this difference is quantified on each plot by the *separation power* $< S^2 >$, defined for a binned distribution as

$$< S^2 >= \sum_{i=1}^{N} \frac{s_i^2 - b_i^2}{s_i + b_i}, \tag{9.3}$$

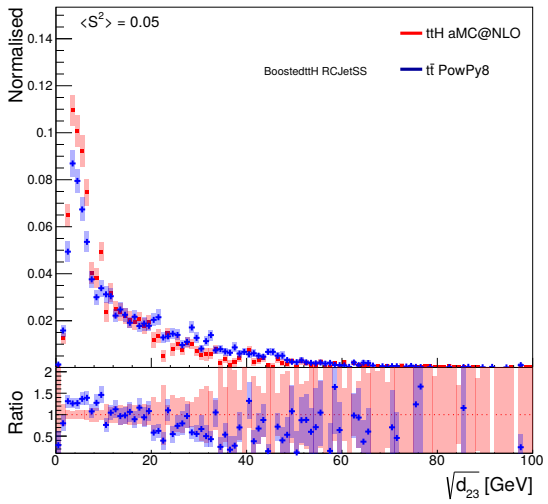
where N is the number of bins in the distribution, s_i is the signal yield in bin i, and b_i is the background yield in bin i, after the total signal and background yields have been normalised to unity. The separation power of the variables used in the classification BDT in the boosted region in Chapter 8 is typically around $< S^2 >\approx 0.01 - 0.08$. The JSS variables all have separation power in this range, and are thus potentially useful as classification BDT variables in this region. However, they are not able to be included in the full analysis. This is because the MC samples currently used in the analysis use the AFII fast detector simulation, in which the calorimeter response is parameterised. This causes a large mismodelling of the calorimeter clusters, making the jet substructure variables unreliable. However, these variables are now available and being used collaboration wide, and the use in $t\bar{t}H, H \to b\bar{b}$ is only a single example application.



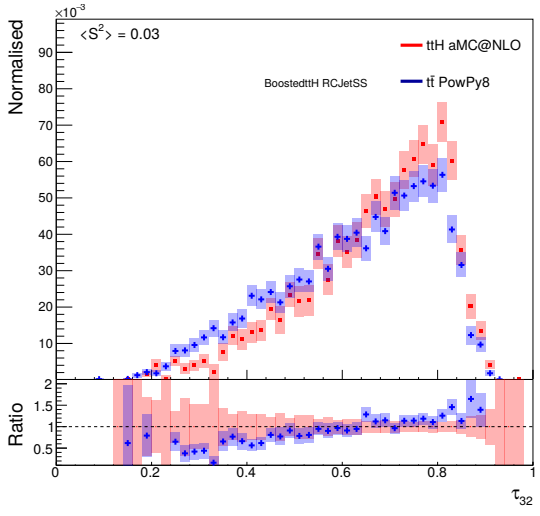
(a)



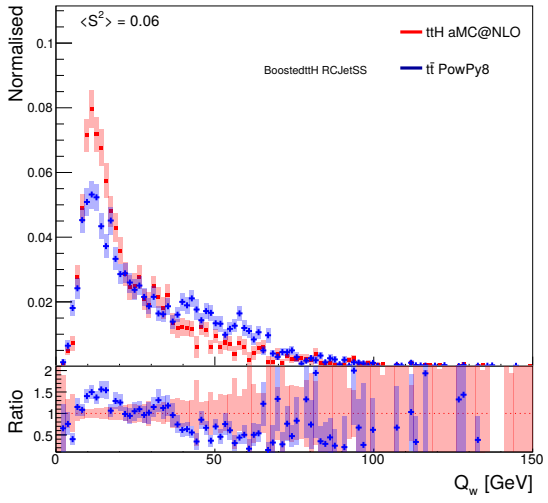
(b)



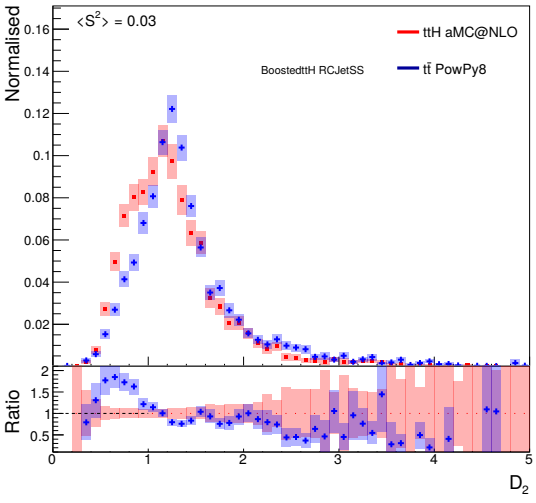
(c)



(d)



(e)



(f)

Figure 9.17: Separation plots for the six reclustered JSS variables, for the Higgs jet in the boosted $t\bar{t}H, H \to b\bar{b}$ selection.

9.3 Multiclass Jet Tagging using Deep Neural Networks

The boosted region in Chapter 8 is defined using requirements on reclustered jets and b-tagging of subjets and additional jets, and this relatively simple selection was able to improve the overall $t\bar{t}H$ analysis after combination with the resolved regions. As discussed in Section 8.3, the reconstructed Higgs jet is truth matched to the true Higgs boson in 47% of the selected events in $t\bar{t}H$ MC. Since the presence of a Higgs boson is the primary difference between signal and background, one clear way to improve the analysis is thus the development of a higher performance jet tagger. Due to the use of reclustered jets in the analysis, existing jet taggers, such as that used to tag top quarks in Chapter 7, are not suitable. Any tagger in use must effectively discriminate between jets produced by top quarks and Higgs bosons, as well as those produced by additional QCD activity. Each of these classes of jet have some similar but distinguishable characteristics: top quarks are the most massive, and contain a b-jet, while Higgs bosons are slightly less massive and contain two b-jets. The third category of jet is expected to be less massive, and contain mostly light quarks / gluons. Thus, a reclustered jet tagger has been developed using a deep neural network (DNN), with a multiclass output corresponding to classification as a Higgs, top, or QCD jet.

9.3.1 DNN Training

The network is trained using the Keras [231] front end to the Tensorflow [232] package. Signal $t\bar{t}H$ MC is used in the training on a jet-by-jet basis, with each reclustered jet of $p_T >$ 200 GeV and mass > 50 GeV, with $\geq$ 2 subjets, included in the training. Jets with both Higgs daughters matched to subjets of the reclustered jet are labelled as Higgs jets, jets matched with the top b-quark daughter and ≥ 1 W-boson daughter are labelled as top jets, and all other jets are labelled as QCD jets. This results in an unbalanced training dataset, and so class weights are applied to the dataset in order to normalise each class of jet to have the same number of training events. The network takes as input 17 variables related to the reclustered jet and its subjets, as summarised in Table 9.3. These variables were selected from a larger set of variables based on the separation powers for each signal and background hypothesis, as defined in Equation 9.3 (see Section 9.2). The least separating variables were removed iteratively and the performance, measured by the diagonality of the confusion matrix, was re-evaluated. The confusion matrix shows both the truth label (y-axis) and predicted label (x-axis) for each jet, and greater diagonality corresponds to higher tagging performance. As the primary difference between signal and background is in the Higgs boson, discrimination of Higgs jets was prioritised. If no significant performance drop was observed the variables were removed, leading to the 17 variables selected.

Table 9.3: List of variables included in DNN training. The JSS variables $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, and Q_W used are calculated using the subjet clustering history, rather than the cluster based approach discussed in Section 9.2. The definition of these variables is discussed in Section 5.3.5.

Variable	Description
m^{rcjet}	Mass of reclustered jet
$\sqrt{d_{12}}$	First splitting scale, calculated using subjets
$\sqrt{d_{23}}$	Second splitting scale, calculated using subjets
Q_W	Minimum invariant mass of subjet pairs, calculated using subjets
N_{subjets}	Number of subjets in reclustered jet
p_T^{subjet1}	p_T of subjet leading in MV2 score
p_T^{subjet2}	p_T of subjet sub-leading in MV2 score
$\text{MV2}^{\text{subjet1}}$	MV2 score of subjet leading in MV2 score
$\text{MV2}^{\text{subjet2}}$	MV2 score of subjet sub-leading in MV2 score
$\Delta R(\text{subjet1}, \text{subjet2})$	Angular separation ΔR of leading and sub-leading in MV2 subjets
$m^{b-\text{jets}}$	Invariant mass of all b-tagged subjets
$m^{\text{light}-\text{jets}}$	Invariant mass of all un-tagged subjets
MV2_{min}	Minimum subjet MV2 score
MV2_{max}	Maximum subjet MV2 score
$\Delta R(\text{subjets})_{\text{min}}$	Minimum ΔR between two subjets
$\Delta R(\text{subjets})_{\text{max}}$	Maximum ΔR between two subjets
MV2^{rest}	MV2 of all subjets except leading and sub-leading in MV2 score

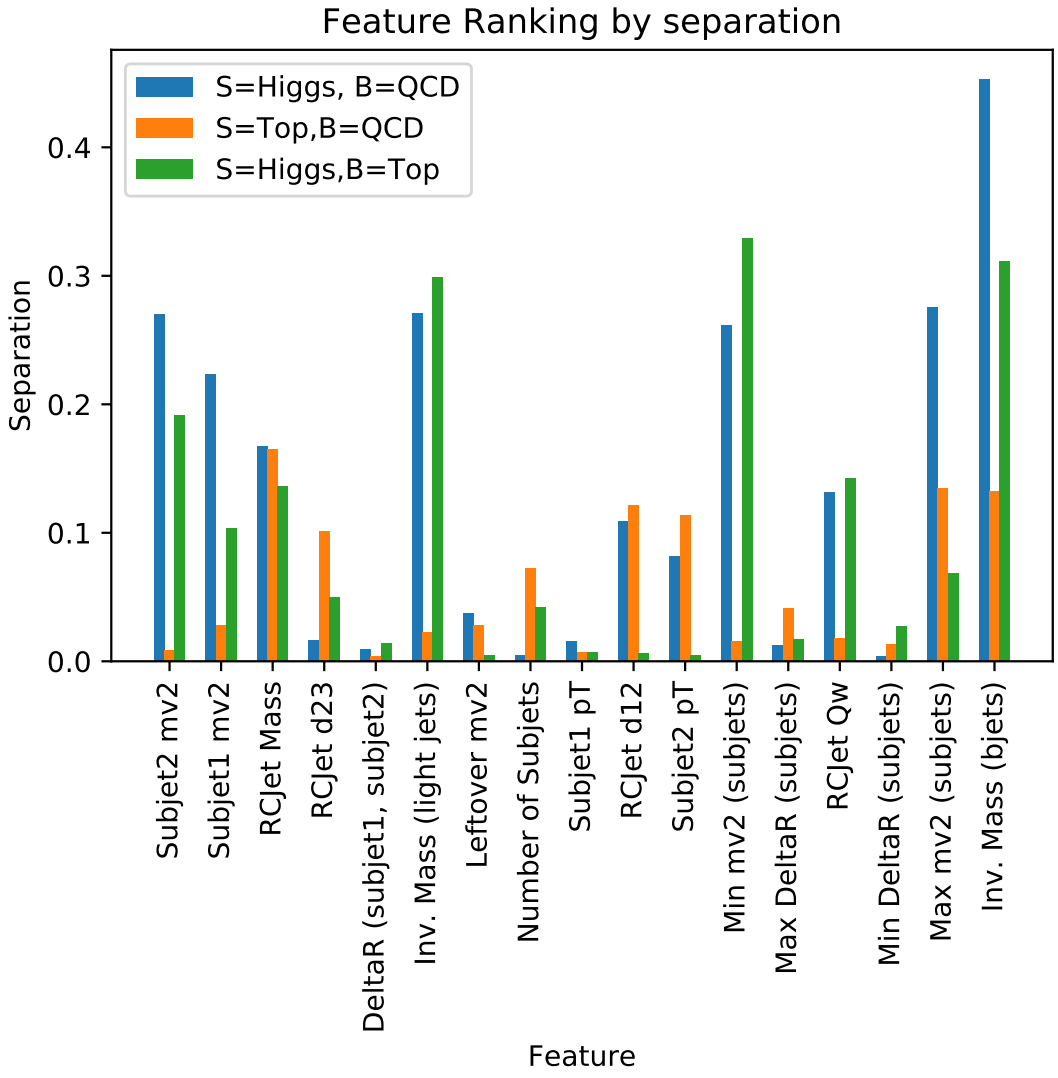


Figure 9.18: The separation, as defined in Equation 9.3, for each input variable used in the DNN training, between Higgs/QCD (blue), Top/QCD (orange), and Higgs/Top (green).

The separation power for these 17 variables are shown in Figure 9.18, and the distributions of three of these variables are shown in Figure 9.19. The invariant mass of the b-tagged subjets, with a clear peak at the Higgs boson mass, is the most separating variable for discrimination of Higgs jets from QCD jets, and also highly ranked for Higgs vs top jet discrimination. The most difficult discrimination in the network is for top vs QCD jets, where the highest ranked variable is the reclustered jet mass, though it shows three distinct peaks as expected. The p_T of the sub-leading subjet discriminates well between top and QCD jets, since QCD jets typically have only a single hard prong. The final confusion matrix is shown in Figure 9.20, demonstrating that Higgs jets are correctly identified 76% of the time and top jets are correctly identified 67% of the time.

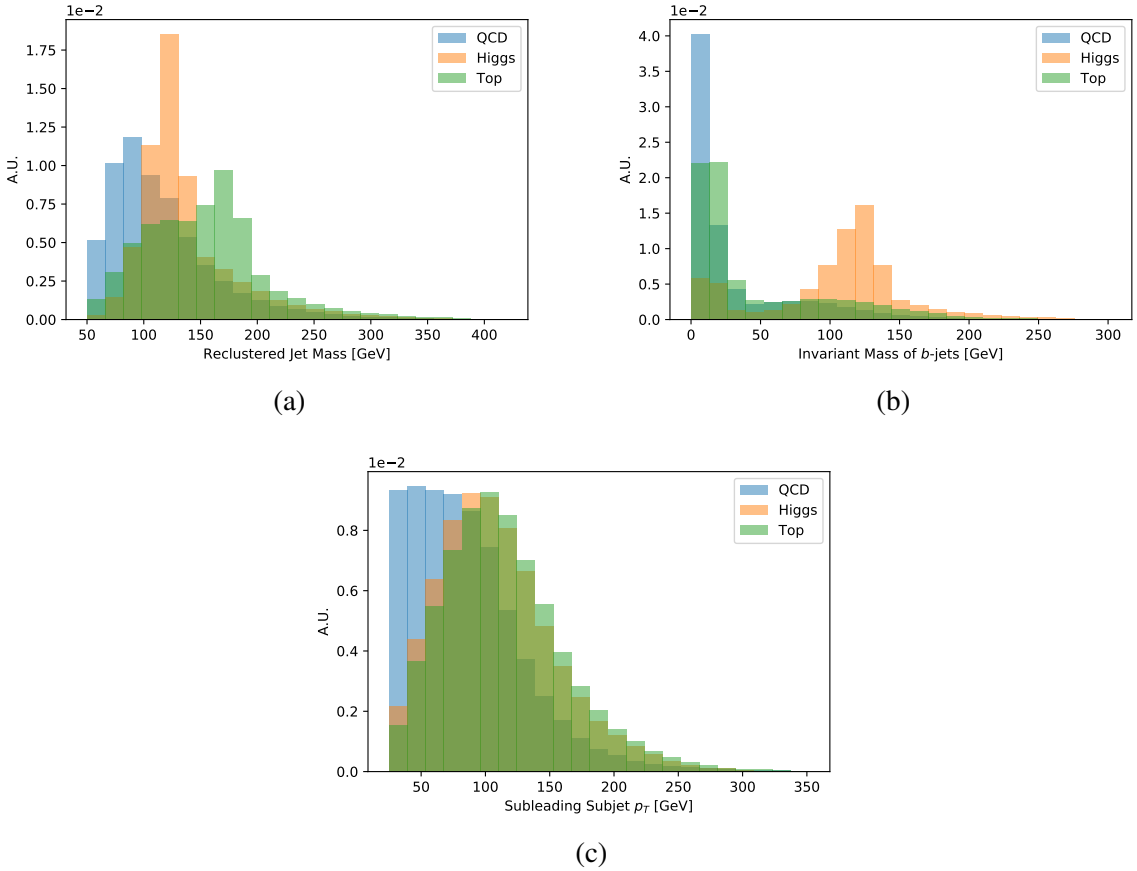


Figure 9.19: Three of the input variables for the DNN. The reclustered jet mass, shown in (a), shows three distinct peaks and thus provides good discrimination in all cases. The invariant mass of the b-tagged subjets, shown in (b), shows a clear peak near the Higgs boson mass for Higgs jets, while both top jets and QCD jets peak at very low values. The sub-leading (in MV2) subjet p_T, shown in (c), is one of few variables which performs best for top vs QCD discrimination, consistent with the hypothesis that most top jets have at least two hard prongs while QCD jets typically have only one.

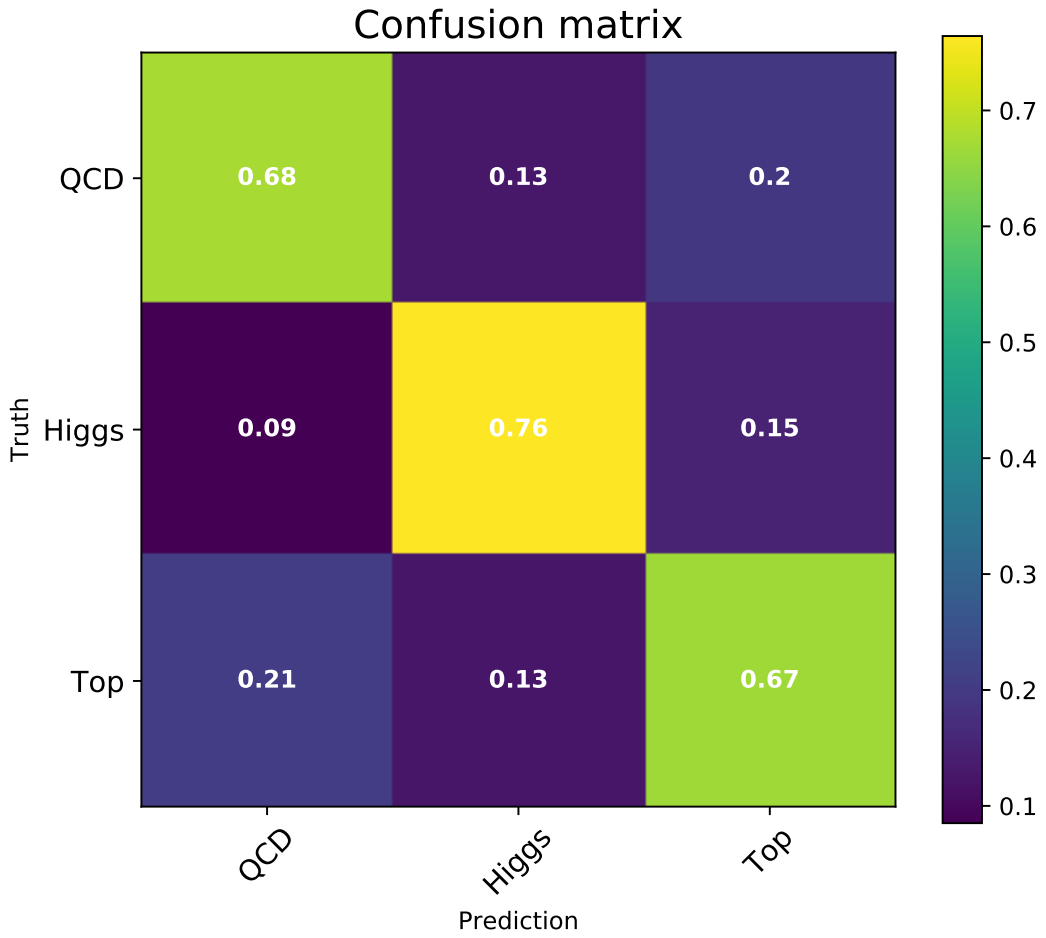
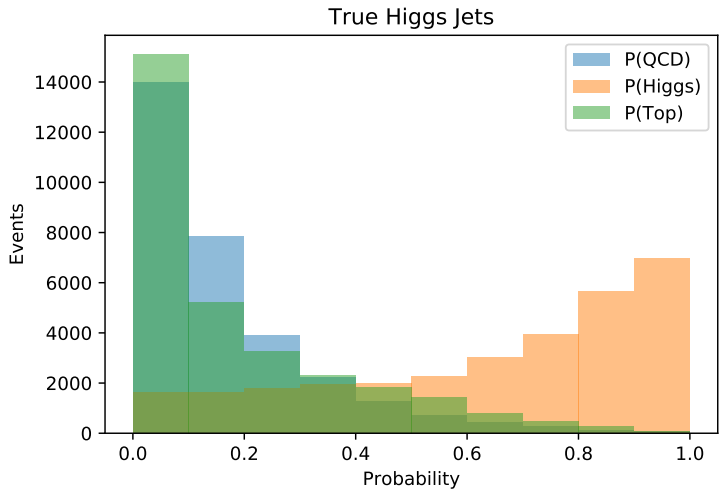


Figure 9.20: Confusion matrix for the DNN output. The x-axis shows the prediction according to the DNN output, while the y-axis shows the truth classification.

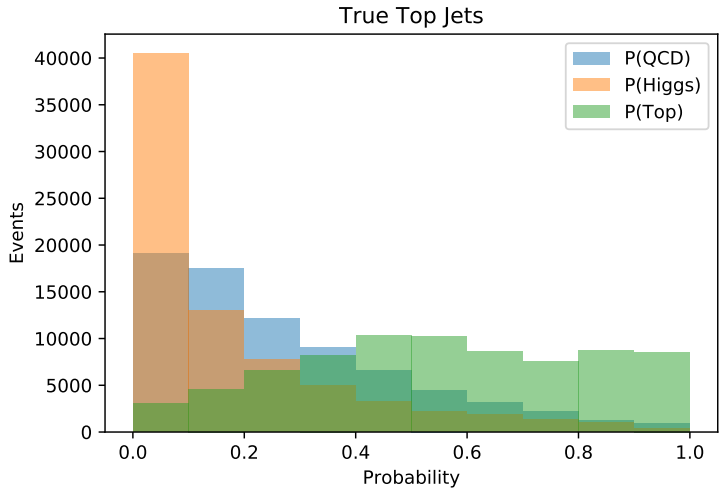
Table 9.4: The hyperparameters used in the DNN training. Any hyperparameters not listed are left to Keras/Tensorflow defaults [231, 232].

Hyperparameter	Value
Number of Layers	3
Nodes per Layer	100
Activation per Layer	ReLU
Learning Rate	0.01
Epochs	50
Decay Rate	10^{-6}
Momentum	0.2

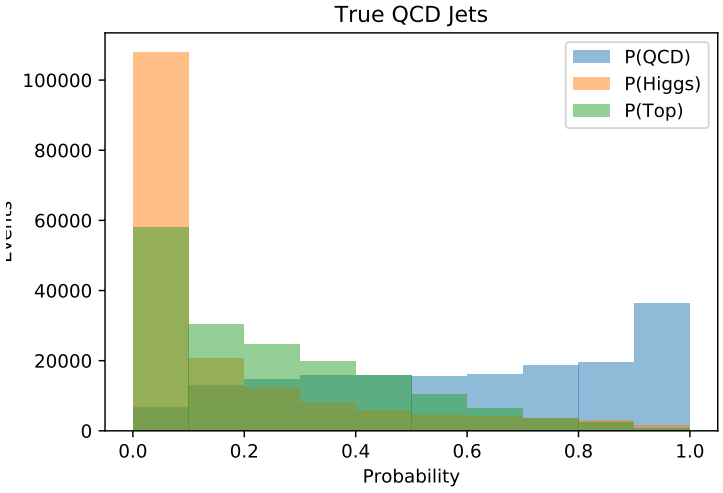
The network hyperparameters are summarised in Table 9.4. The network architecture is chosen to be 3 layers of 100 nodes, having been chosen to balance performance and training time – tests with larger number of layers or nodes did not significantly improve performance while taking significantly longer to train. Similarly, the number of epochs was selected to be 50 to prevent overtraining. The hidden layers each use ReLU activation. The choice of each hidden layer to have the same number of nodes and the same activation is not a requirement, but is chosen for simplicity and not optimised. The final layer is a 3 node softmax output, which constrains the output nodes to sum to unity. This corresponds to probabilities for the three possible labels: P_H, P_t and P_Q for Higgs, top and QCD respectively, with $P_H + P_t + P_Q = 1$. In this way, the jet can be tagged by assigning it to the category which has the highest output score. The other hyperparameters of the network were set to a reasonable default and were not optimised. The output distributions from the DNN are shown in Figure 9.21, showing good separation for each category of jet.



(a)



(b)



(c)

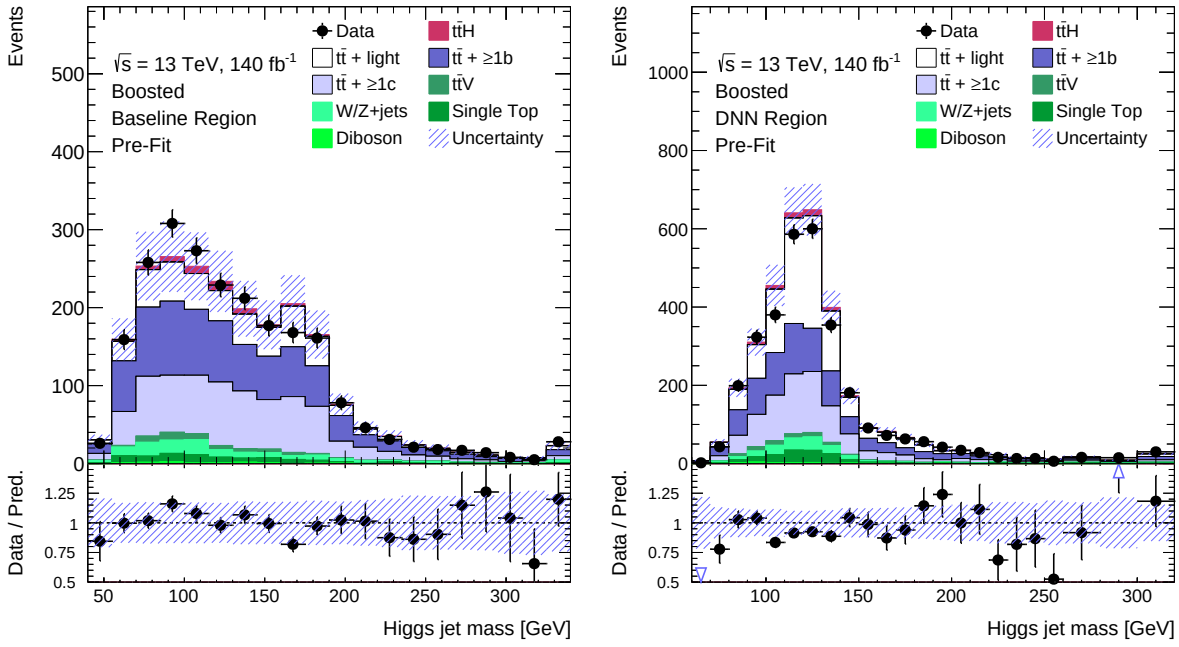
Figure 9.21: DNN output distributions, for truth matched (a) Higgs jets, (b) top jets, and (c) QCD jets. In each case, the three DNN outputs are shown, with the output corresponding to the truth matched class clearly separated from the others, demonstrating good performance.

Table 9.5: Yields in the baseline and DNN selections, based on 140 fb^{-1} of MC.

Process	Baseline	DNN Selection
$t\bar{t}H$	56.0 ± 7.5	74.1 ± 7.0
$t\bar{t}$+light	424.8 ± 166.0	1216.3 ± 272.2
$t\bar{t}$+b	758.0 ± 88.3	835.3 ± 72.0
$t\bar{t}$+c	707.0 ± 120.8	834.3 ± 88.3
$t\bar{t}W$	17.0 ± 3.5	30.8 ± 4.6
$t\bar{t}Z$	38.8 ± 6.2	50.5 ± 6.8
Z+jets	11.0 ± 5.3	16.5 ± 7.0
W+jets	110.7 ± 59.5	147.2 ± 67.0
Single Top	74.3 ± 14.2	162.3 ± 17.1
Diboson	11.3 ± 6.7	18.6 ± 10.1
Total	2208.8 ± 379.3	3385.8 ± 345.4
S / B	2.6%	2.2%
$S/\sqrt{B}$	1.21	1.28

9.3.2 Selecting Events using DNN Tagger

With the DNN trained, events are selected requiring at least one lepton with $p_T > 27$ GeV passing single lepton triggers, and two reclustered jets each with $p_T > 200$ GeV, $m > 50$ GeV, and at least two subjets. One of these reclustered jets must be tagged as a Higgs jet according to the DNN output, by having $P_H > P_t$ and $P_H > P_Q$. Similarly, a second reclustered jet must be tagged as a top jet by having $P_t > P_H$ and $P_t > P_Q$. In the case that there is more than one Higgs or top jet, the jet with the largest P_H or P_t respectively is chosen. The signal and backgrounds are estimated from MC samples as described in Section 6.8. The expected yields in the DNN region, corresponding to 140 fb^{-1} of $\sqrt{s} = 13$ TeV data, are compared to the existing boosted region from Chapter 8, referred to here as the *baseline* region, in Table 9.5. The DNN region has a higher yield in both signal and background than the baseline, with the increase in background yield primarily from the $t\bar{t}$+light background. The $S/\sqrt{B}$ increases from 1.21 in the baseline region to 1.28 in the DNN region despite this. Furthermore, the fraction of correctly matched Higgs jets in the DNN region is 56.4%, compared to 45.8% in the baseline region. The Higgs jet mass distributions, shown in Figure 9.22, show a much broader distribution in the baseline region which selects the Higgs based on b-tagging information only, including a secondary peak around the top mass. The narrower peak in the DNN region is expected, since the jet mass is used as an input to the DNN tagger.

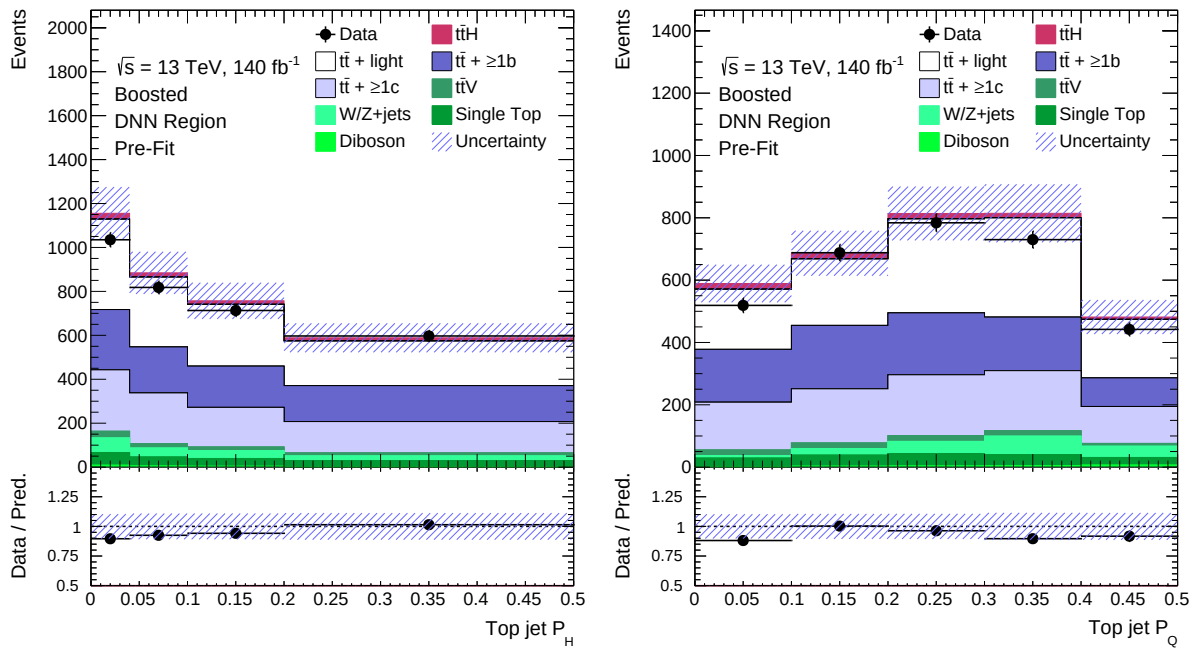


(a) (b)

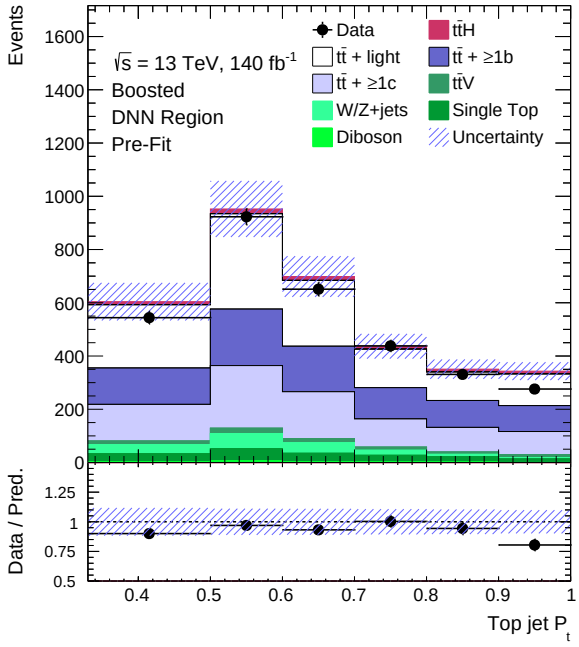
Figure 9.22: The Higgs jet mass distribution, in (a) the baseline region and (b) the DNN region. The uncertainty band corresponds to the instrumental systematic uncertainties, with the modelling systematic uncertainties excluded.

To check the performance of the DNN in data, the DNN output distributions for the selected top jet are shown in Figure 9.23. The top jets in these distributions are correctly matched to the hadronic top quark in 61.1% of events in inclusive $t\bar{t}$ MC, with the rest truth matched to QCD jets. These distributions are a good check of the modelling of the DNN in the signal region, since they are not sensitive to the signal. They show good agreement between data and MC within the shown systematic uncertainty, which includes the instrumental uncertainties but not the modelling uncertainties. The P_H distribution peaks at zero, as expected. The P_t distribution peaks just above 0.5, rather than near 1.0, while the P_Q distribution is quite flat between 0.0 and 0.5. Comparing to Figure 9.21, these shapes make sense given the fraction of events in which the top jets are correctly matched. Figure 9.24 show the kinematics of the top jets. There is a slight hint of the familiar top p_T slope observed in Chapter 7, as well as a slight slope in the mass distribution, though this is mostly covered by the uncertainty band even without the modelling uncertainties included. Good agreement is observed between data and MC for the top jet pseudo-rapidity.

Figures 9.25 and 9.26 show the $t\bar{t}H$ signal yields in each bin of the pseudo-continuous b-tagging (PCB) score of the four small jets leading in PCB score in the event, for the DNN region and baseline region respectively. This classification, which was used in the resolved analysis in Chapter 8, uses a score based on the tightest working point at which the jets are b-tagged, with 60%=5, 70%=4, 77=3, 85=2, and 100%=1, and subsequently categorises on this basis. Those

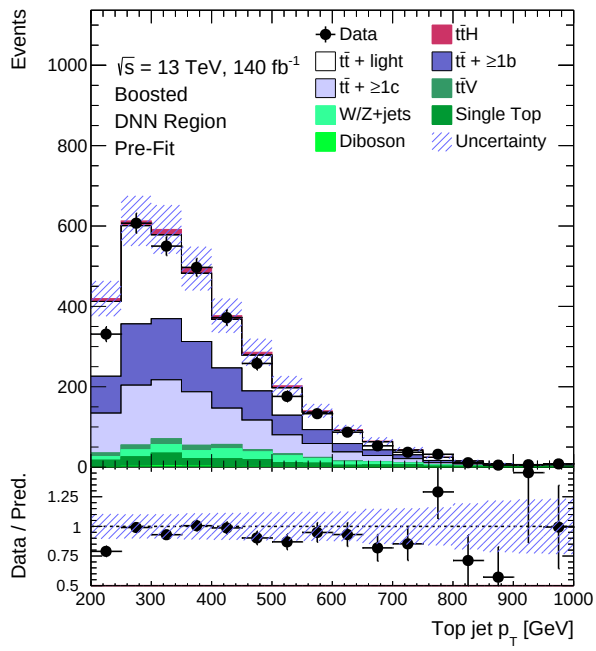


(a) (b)

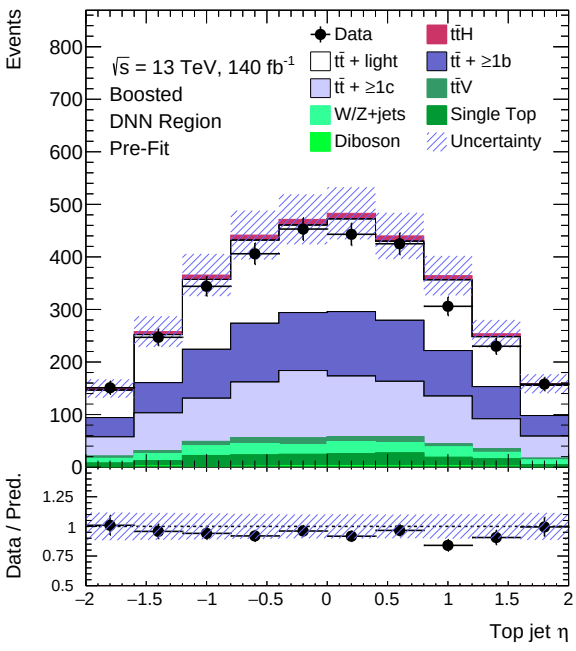


(c)

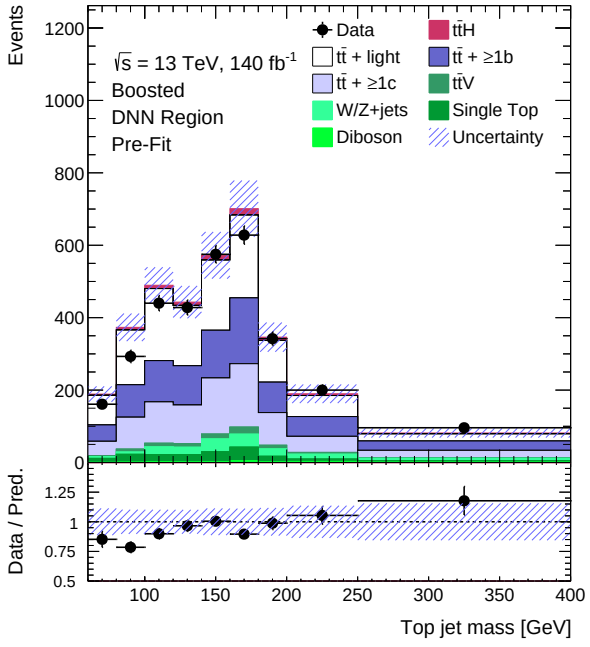
Figure 9.23: The output DNN distributions for the selected top jet in the DNN region. The uncertainty band corresponds to the instrumental systematic uncertainties, with the modelling uncertainties excluded.



(a)

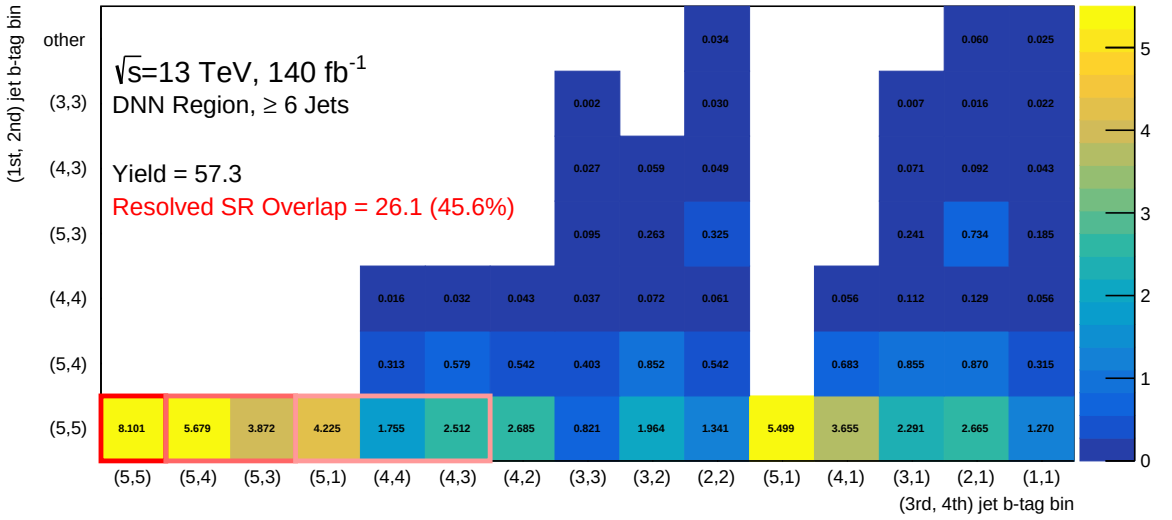


(b)

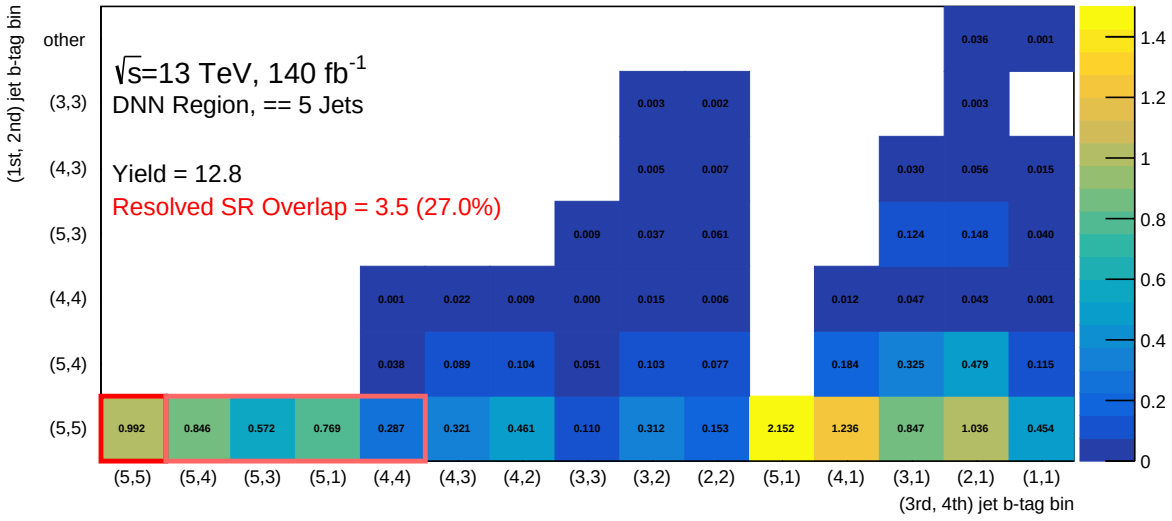


(c)

Figure 9.24: Kinematic distributions for the selected top jet in the DNN region. The uncertainty band corresponds to the instrumental systematic uncertainties, with the modelling uncertainties excluded.

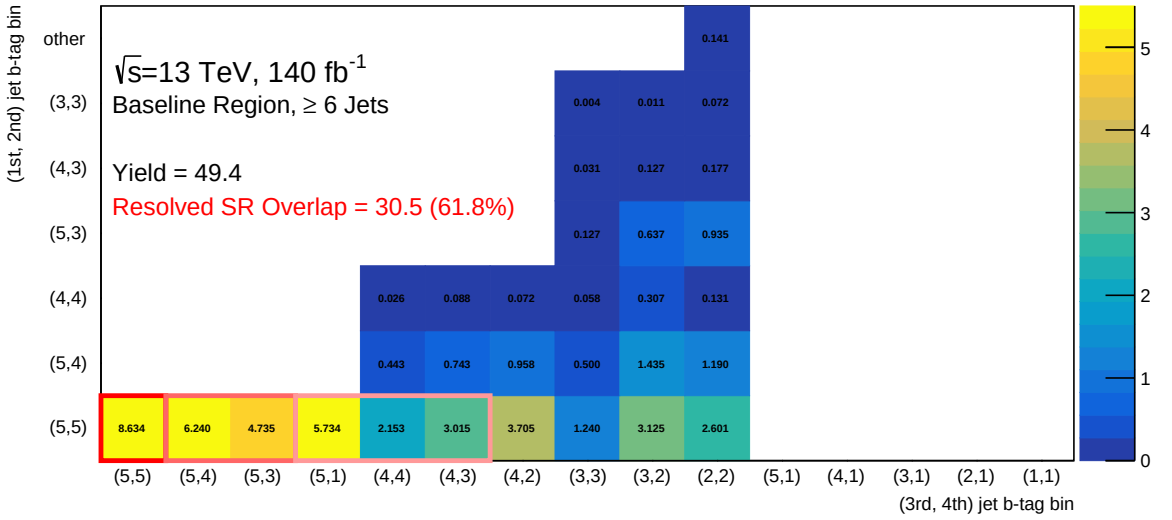


(a)

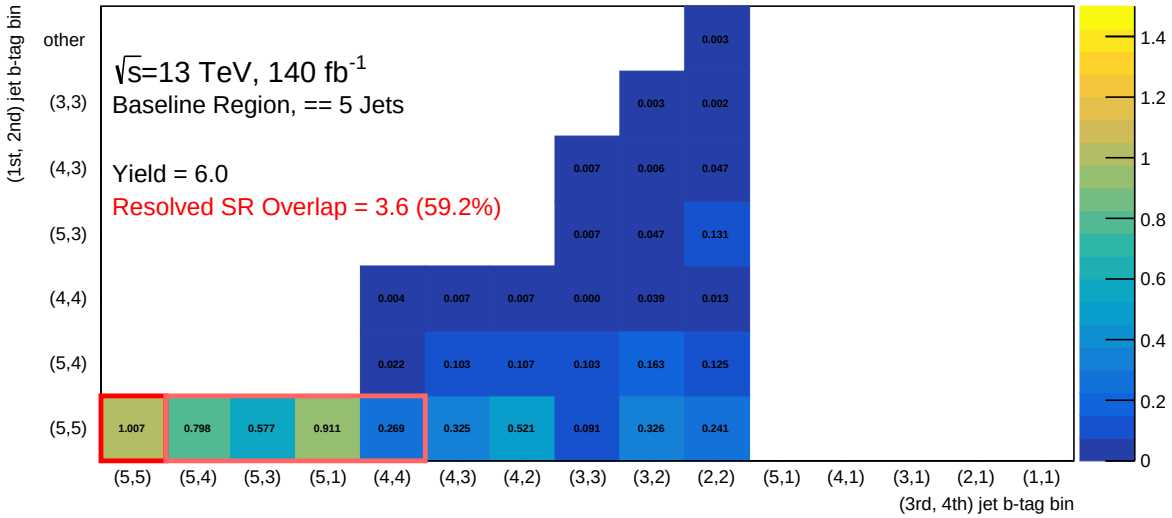


(b)

Figure 9.25: The overlap with the resolved regions, highlighted in red, in the DNN region, for (a) ≥6 jet events, and (b) 5 jet events. The y-axis represents the PCB bin for the leading and sub-leading b-jet, and the x-axis represents the PCB bin for the third and fourth b-jet, in the same manner as Figures 8.2 and 8.1. The z-axis represents the signal yield. Bins which are highlighted in different shades of red represent the different signal regions of the resolved analysis as defined in Chapter 8. An additional 4.4 events are expected in 4 jet events in the DNN region.



(a)



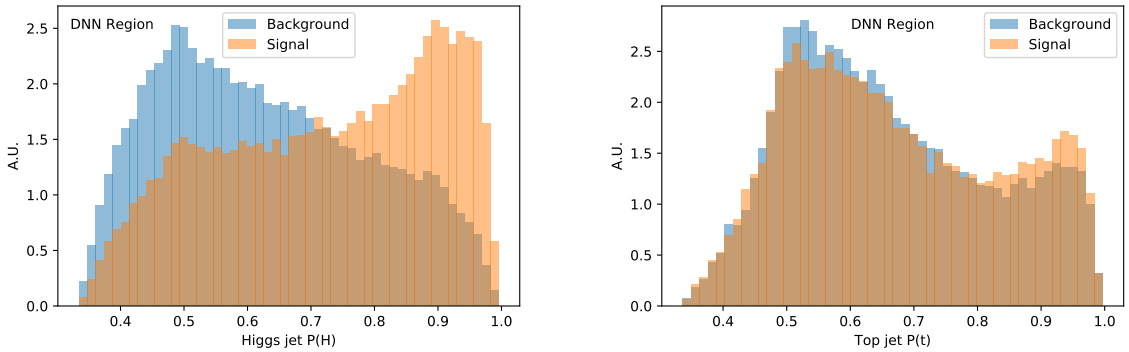
(b)

Figure 9.26: The overlap with the resolved regions, highlighted in red, in the baseline region, for (a) ≥6 jet events, and (b) 5 jet events. The y-axis represents the PCB bin for the leading and sub-leading b-jet, and the x-axis represents the PCB bin for the third and fourth b-jet, in the same manner as Figures 8.2 and 8.1. The z-axis represents the signal yield. Bins which are highlighted in different shades of red represent the different signal regions of the resolved analysis as defined in Chapter 8.

bins which are highlighted in red represent the signal regions of the resolved analysis as defined in Chapter 8. The overlap with the signal regions of the resolved analysis is reduced in the DNN region, with 40% of DNN region events coming from resolved signal regions, compared to 62% in the case of the baseline region. This will mean that the relative performance of the DNN region compared to the baseline region is likely to improve in the full combination with the resolved regions, when the orthogonality of regions is important and boosted events are vetoed from resolved regions.

9.3.3 Classifier Training in DNN Region

Starting from the selection based on the DNN, a new classification BDT is trained to separate signal and background in this region. The input variables used in the training of the classification BDT in the DNN region are summarised in Table 9.6. These variables were chosen in a similar way to the DNN inputs, starting with a larger pool of variables and iteratively removing the lowest ranked until performance begins to drop. Most of the variables are generally similar to those used in Chapter 8, shown in Figure 8.1. In addition, some of the DNN outputs themselves are included in the BDT, since these contain further useful information beyond the initial event selection. For example, the Higgs and top jet DNN scores, shown in Figure 9.27, are included. In particular, the Higgs jet P_H has very good separation between signal and background, since in the background the Higgs jet is always a mistag. The BDT is trained using the XGBoost package [233], with hyperparameters as shown in Table 9.7. As with the DNN, these hyperparameters have not been formally optimised, and were mostly chosen to be as similar as possible to the TMVA hyperparameters used in Chapter 8. The signal and background separation for the DNN region classification BDT is shown in Figure 9.28, and is compared to the baseline region classifiers in Section 9.3.4.



(a) (b)

Figure 9.27: The (a) Higgs jet Higgs score, and (b) top jet top score, used in the DNN region classification BDT training. Signal $t\bar{t}H$ is shown in orange, with the sum of all backgrounds shown in blue.

Table 9.6: Input variables used for the classification BDT in the DNN region.

Variable	Description
$\text{MV2}^{\text{Higgs}}$	Total MV2 score of the Higgs subjets
MV2^{top}	Total MV2 score of the top subjets
MV2^{add}	Total MV2 score of the additional small jets
m_{Higgs}	Higgs jet mass
E_T^{miss}	Missing transverse energy in the event
N_{jets}	Number of small jets in the event
$\Delta R(\text{Higgs, top})$	Angular separation between the Higgs and Top jets
$\Delta R(\text{top, lepton})$	Angular separation between the Top jet and the lepton
$\Delta R(\text{Higgs, lepton})$	Angular separation between the Higgs jet and the lepton
P_H^{Higgs}	DNN Higgs score of the Higgs jet
P_t^{top}	DNN Top score of the Top jet
$\Delta P_t(\text{top, Higgs})$	Difference in Top score for the Top and Higgs jets

Table 9.7: Hyperparameters for the classification BDT training in the DNN region. No formal optimisation procedure was performed.

Hyperparameter	Value
Learning Rate	0.01
Max. Depth	3
Number of Trees	700
Min. Child Weight	1.0

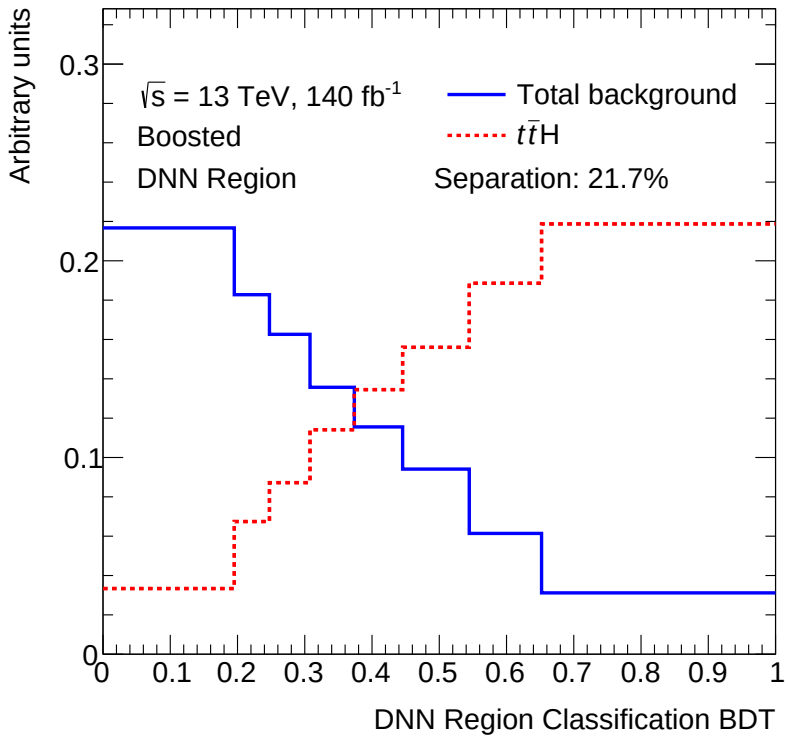


Figure 9.28: Separation plot for the classification BDT in the DNN region.

9.3.4 Classifier Training in Baseline Region

An alternative application of the DNN is to incorporate the DNN scores into the classification BDT in the baseline region from Chapter 8, where Higgs and top candidates are selected using subjet *b*-tagging only. As with the DNN region, the Higgs and top scores of both the Higgs and top jets, shown in Figure 9.29, contain separating information between signal and background in this region which can be included in the final classifier to improve performance. Two new BDTs are trained, using the same inputs listed in Table 8.1, with one also incorporating the additional four variables from the DNN shown in Figure 9.29. Fits to these BDTs with and without the DNN variables are referred to as the *retrained baseline* and *baseline+DNN* fits respectively. As with the DNN region, they are trained with the XGBoost package [233] using hyperparameters as similar to those from Chapter 8 as possible, as listed in Table 9.7. A cross-check, retraining the baseline BDT using TMVA, finds very similar performance between the two packages. The separation plots are shown in Figure 9.30, showing the increased separation obtained when including the DNN scores in the training. The separation in the DNN region BDT, shown in Figure 9.28, is higher than both of the baseline BDTs.

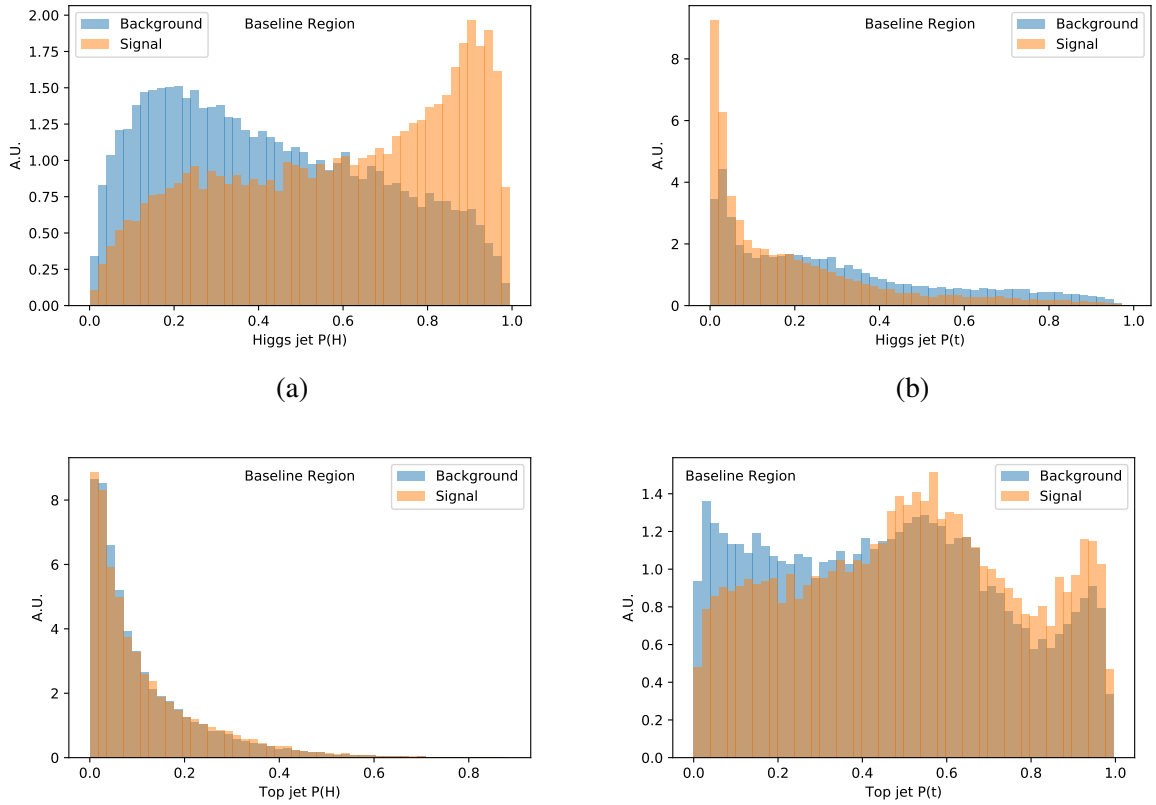
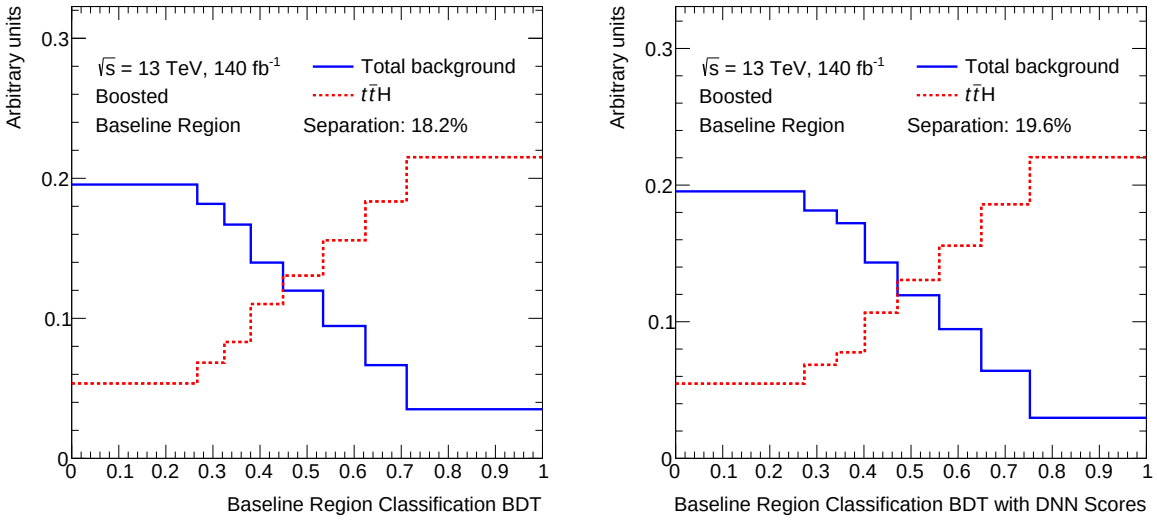


Figure 9.29: The DNN (left) Higgs and (right) top scores for the (top) Higgs candidate, and (bottom) top candidate, in the baseline region. Signal $t\bar{t}H$ is shown in orange, with the sum of all backgrounds in blue, demonstrating the separation power these bring to the classification BDT in this region.

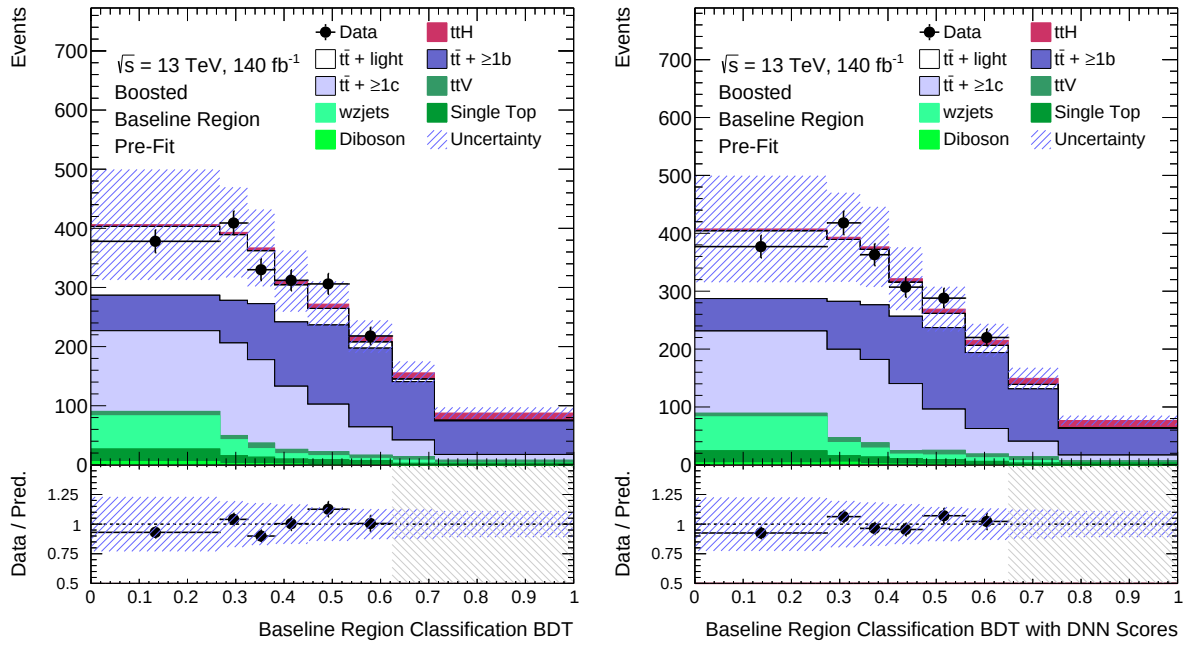


(a) (b)

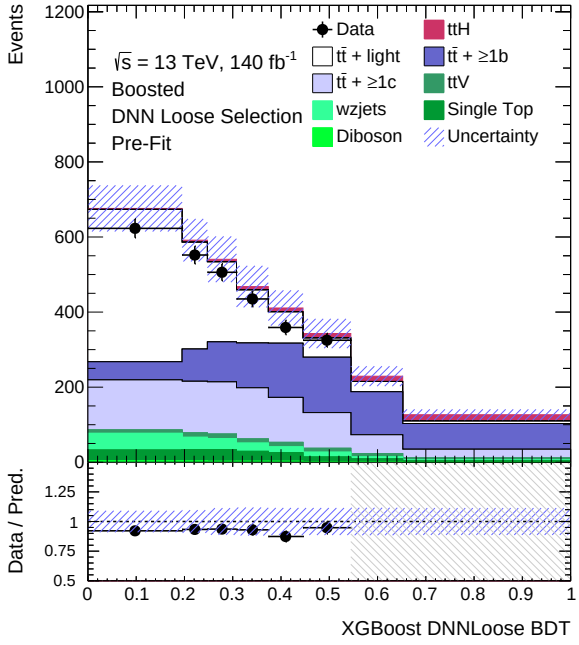
Figure 9.30: Separation plots for the (a) retrained baseline, and (b) baseline+DNN regions. The separation power, as defined in Equation 9.3, is shown on the plot, demonstrating the increased separation power obtained by adding the DNN variables.

9.3.5 Fits

In order to evaluate the performance of the new region and classifier options, Asimov fits are performed to the three options: the DNN selection and classification BDT described in Section 9.3.3, the baseline region with the retrained classification BDT with inputs as from Chapter 8, and the baseline region with the classification BDT including the DNN output scores, described in Section 9.3.4. The data and MC distributions are shown in Figure 9.31, with any bin containing more than 5% signal events blinded in the data. In the bins shown, the data and MC agreement is good, though there is a slight overall overestimation in the DNN region. Since the fits are to the boosted regions only, the background normalisation factors are fixed to the values found in the analysis of Chapter 8, where $k_b = 1.24$ and $k_c = 1.63$, due to the lack of constraining power without the inclusion of the resolved control regions. This choice was cross-checked by fixing these to alternative values, such as $k_c = k_c = 1.0$ or $k_b = 1.63$ and $k_c = 1.24$, and the conclusions were consistent results regardless of which values were chosen. Three fit models were tested in each region, each using Asimov data corresponding to 140 fb^{-1}: statistics only, statistics + instrumental systematics only, and full fits including both instrumental and modelling systematics. The instrumental only fits, which include the systematics due to the objects used such as jets, b-tagging, leptons, and triggers – were performed due to observed instability in the fits caused by a lack of MC statistics in the alternative samples used to define the modelling uncertainties. It is these modelling systematics that were the dominant ones in the results of Chapter 8, in particular for the $t\bar{t}$ + $b\bar{b}$ background. The instrumental systematics are as in



(a) (b)



(c)

Figure 9.31: The data and MC distributions, for (a) the retrained baseline BDT, (b) the baseline+DNN BDT, and (c) the DNN region BDT. Bins which contain more than 5% signal in MC are blinded in data. The hashed area represents the systematic uncertainties including only the instrumental systematics.

Chapter 8, though with updated calibrations and scale factors for the full Run 2 dataset. The preliminary set of modelling systematics for the $t\bar{t}$ background included in the full fits are:

- Matrix Element Generator: replacing the nominal Powheg + Pythia8 $t\bar{t}$ signal sample with an aMC@NLO + Pythia8 sample and re-evaluating the background distribution, taking the difference between the two as the uncertainty. This is performed separately for each of the three components ($t\bar{t}$+≥1b, $t\bar{t}$+≥1c, $t\bar{t}$+light) of the $t\bar{t}$ background and treated as uncorrelated.

- Parton Shower and Hadronisation: replacing the nominal Powheg + Pythia8 $t\bar{t}$ signal sample with a Powheg + Herwig7 sample and re-evaluating the background distribution, taking the difference between the two as the uncertainty. This is performed separately for each of the three components ($t\bar{t}$+≥1b, $t\bar{t}$+≥1c, $t\bar{t}$+light) of the $t\bar{t}$ background and treated as uncorrelated.

- Initial State Radiation: applying an additional event weight that reproduces the effects of varying the Powheg and Pythia8 settings and tunes related to the amount of initial state radiation, and repeating the fit. This is applied separately for each of the three components ($t\bar{t}$+≥1b, $t\bar{t}$+≥1c, $t\bar{t}$+light) of the $t\bar{t}$ background and treated as uncorrelated.

Similarly, uncertainties on the modelling of the $t\bar{t}H$ signal are also included. Those are:

- Matrix Element Generator: replacing the nominal Powheg + Pythia8 $t\bar{t}H$ signal sample with an aMC@NLO + Pythia8 sample and re-evaluating the signal distribution, taking the difference between the two as the uncertainty.

- Parton Shower and Hadronisation: separately replacing the nominal Powheg + Pythia8 $t\bar{t}H$ signal sample with an aMC@NLO + Pythia8 sample and with an aMC@NLO + Herwig7 sample, and re-evaluating the signal distribution in both cases. The difference between the two alternative samples is taken as the uncertainty.

Plots demonstrating the effect of the systematic on the matrix element calculation on the BDT output distributions are shown in Figure 9.32. The overall difference between the two samples is around 15% in all three distributions, with the difference between each bin of the distribution in the two MC samples shown in dashed red. To minimise the effect of fluctuations, smoothing is applied, which is shown in solid red, with the symmetrised uncertainty in blue. Despite the smoothing, the shape of the uncertainty is still highly sensitive to the fluctuations in the samples, which is particularly clear in Figure 9.32a. For comparison, Figure 9.33 shows the effect of the first light mistagging component of the b-tagging eigenvalues (see Section 5.3.2) on the $t\bar{t}$+light background. This uncertainty is evaluated using event weights corresponding to variation of the light jet mistagging calibration, and as such does not suffer from the problem of low MC statistics.

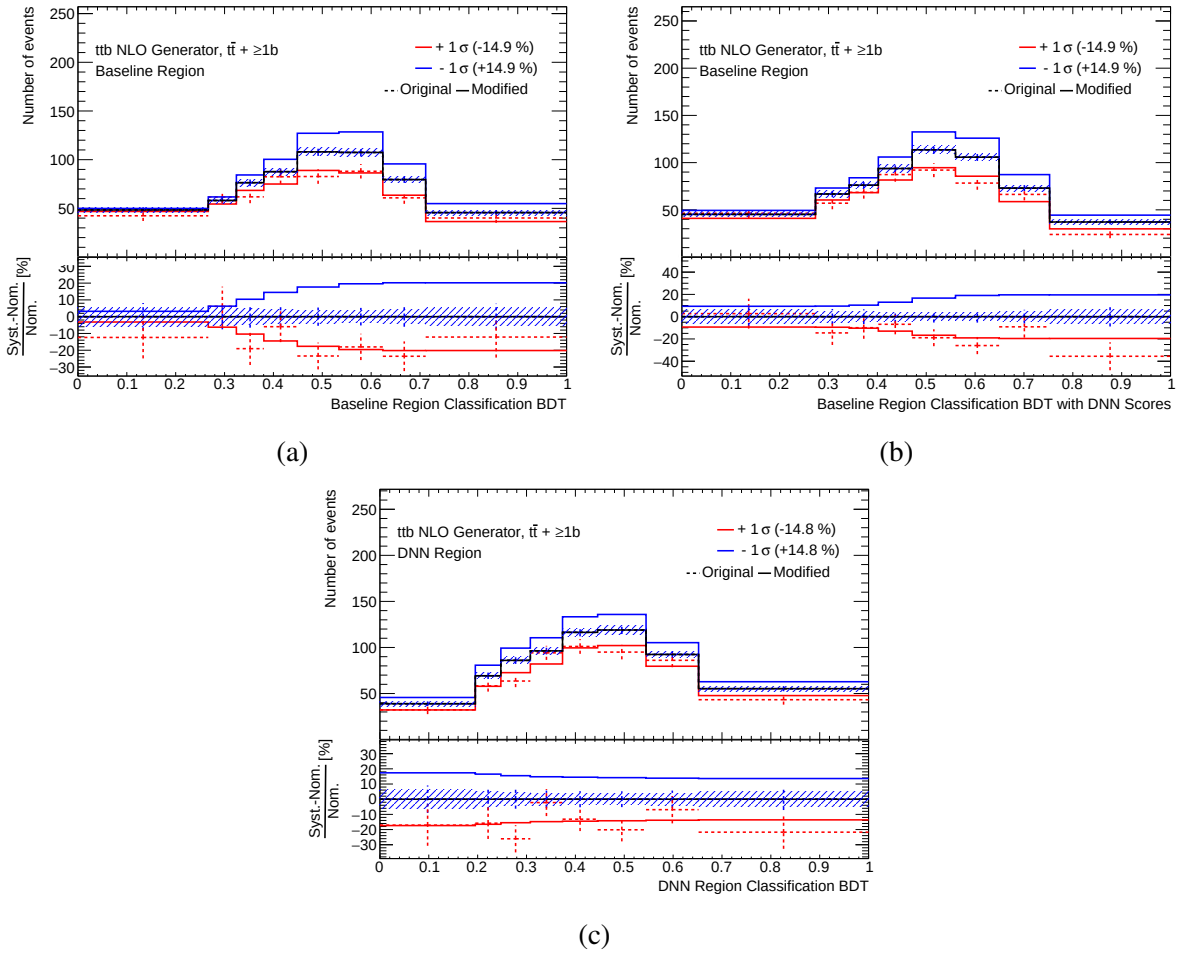
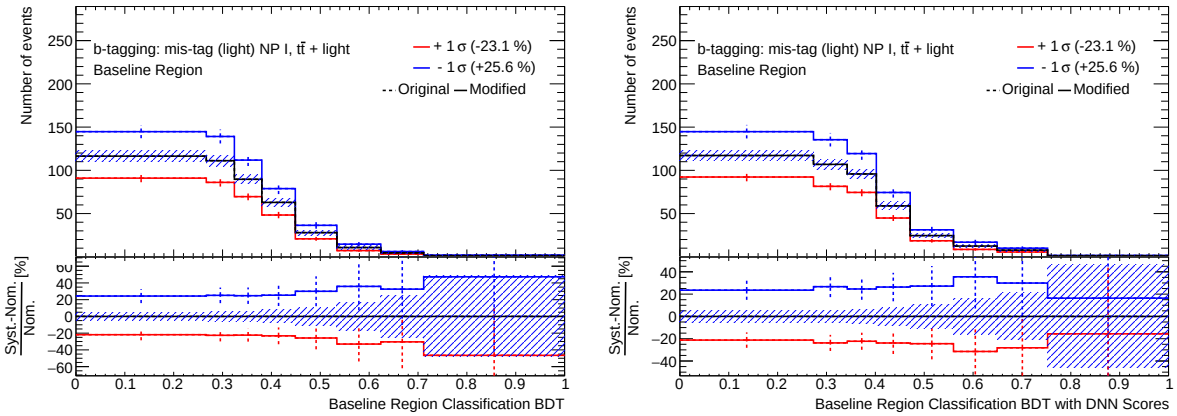
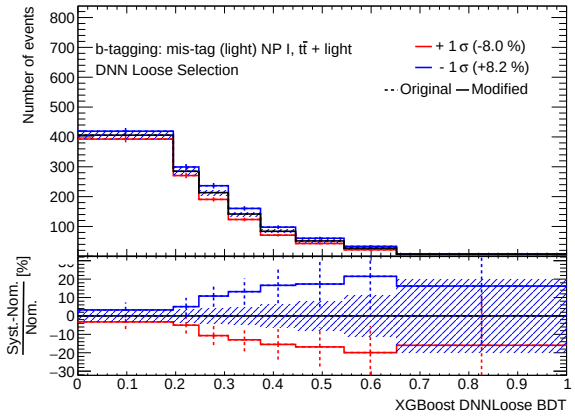


Figure 9.32: Plots demonstrating the effect of the matrix element generator systematic, in (a) the retrained baseline, (b) the baseline+DNN, and (c) the DNN regions. The nominal distribution is shown in black. The up and down variations of the distribution, applied via event weights, is shown in red and blue respectively, with no smoothing applied.



(a) (b)



(c)

Figure 9.33: Plots demonstrating the effect of the first light mistagging component from the *b*-tagging eigenvalues, in (a) the retrained baseline, (b) the baseline+DNN, and (c) the DNN regions. The nominal sample is shown in black. The alternative sample distribution is shown in dashed red, with the smoothed systematic distribution in solid red and symmetrised in solid blue.

Table 9.8: Results from the 140 fb^{-1}, boosted only fits to the three analysis options. The numbers in brackets represent the RMS on the uncertainty band, determined using 1000 toy experiments.

Region	Statistics Only	Instrumental Systematics.	Full Systematics
Baseline	$1.0^{+0.58}_{-0.56}(\pm 0.01)$	$1.0^{+0.93}_{-0.86}(\pm 0.05)$	$1.0^{+1.31}_{-1.26}(\pm 0.13)$
Baseline+DNN	$1.0^{+0.54}_{-0.52}(\pm 0.01)$	$1.0^{+0.85}_{-0.84}(\pm 0.04)$	$1.0^{+1.09}_{-1.02}(\pm 0.11)$
DNN Selection	$1.0^{+0.52}_{-0.50}(\pm 0.01)$	$1.0^{+0.84}_{-0.83}(\pm 0.03)$	$1.0^{+1.20}_{-1.18}(\pm 0.15)$

In order to understand the limitations of the MC statistics in each of these fits, 1000 toy fits were performed in each case, using poisson varied event weights for each process, on both the nominal and systematic MC distributions. The performance of the different fit models is compared by looking at the mean expected error on μ, and the root mean square of the distribution of the error in the toy fits. These distributions are shown in Figure 9.34. The statistics only fits, shown in blue, show narrow Gaussian distributions, as expected. The instrumental fits, shown in purple, are similar, with slightly wider, approximately Gaussian distributions. The full fits, shown in red, have a much wider distribution, demonstrating the large instability of the modelling systematics. In particular, the nominal fit without toy weights, shown as a dashed line for each fit model, lies far from the centre of the distribution for the DNN selection with the full systematics. This emphasises why the toy fits are necessary to truly evaluate the three regions. The results are summarised in Table 9.8.

In the statistics only fit, the selection utilising the DNN to tag the Higgs and top jets performs best, improving on the baseline region by 8%, with $> 3\sigma$ significance. The baseline+DNN region, which has the same event selection as Chapter 8 but includes the DNN variables in the classification BDT, also improves on the baseline region in the statistics only fits, in this case by 4%.

When including the instrumental systematics, the DNN and the baseline+DNN perform similarly, both improving on the baseline by around 6%, though only with around 2σ significance. The pulls and constraints of the 20 largest systematics for these fits are shown in Figure 9.35. The largest two systematics in the fits to the baseline and baseline+DNN BDTs are the uncertainty on the inclusive $t\bar{t}$ cross-section and the first nuisance parameter of the jet energy scale. The largest uncertainty in the DNN selection is the first light mistagging component from the b-tagging eigenvalues, which is unsurprising given the greatly increased $t\bar{t}$+light background. This uncertainty has a 30% correlation to the inclusive $t\bar{t}$ cross-section and a 16% correlation to the first jet energy scale nuisance parameter, which causes these nuisance parameters to have a smaller effect in the DNN selection. Otherwise, the set of systematics appearing in these plots are generally similar.

In the full fits, the statistical error from the toys is around 10%, larger than the improvement observed in the statistics only fits. The expectation is that the effect of the systematic uncertainties

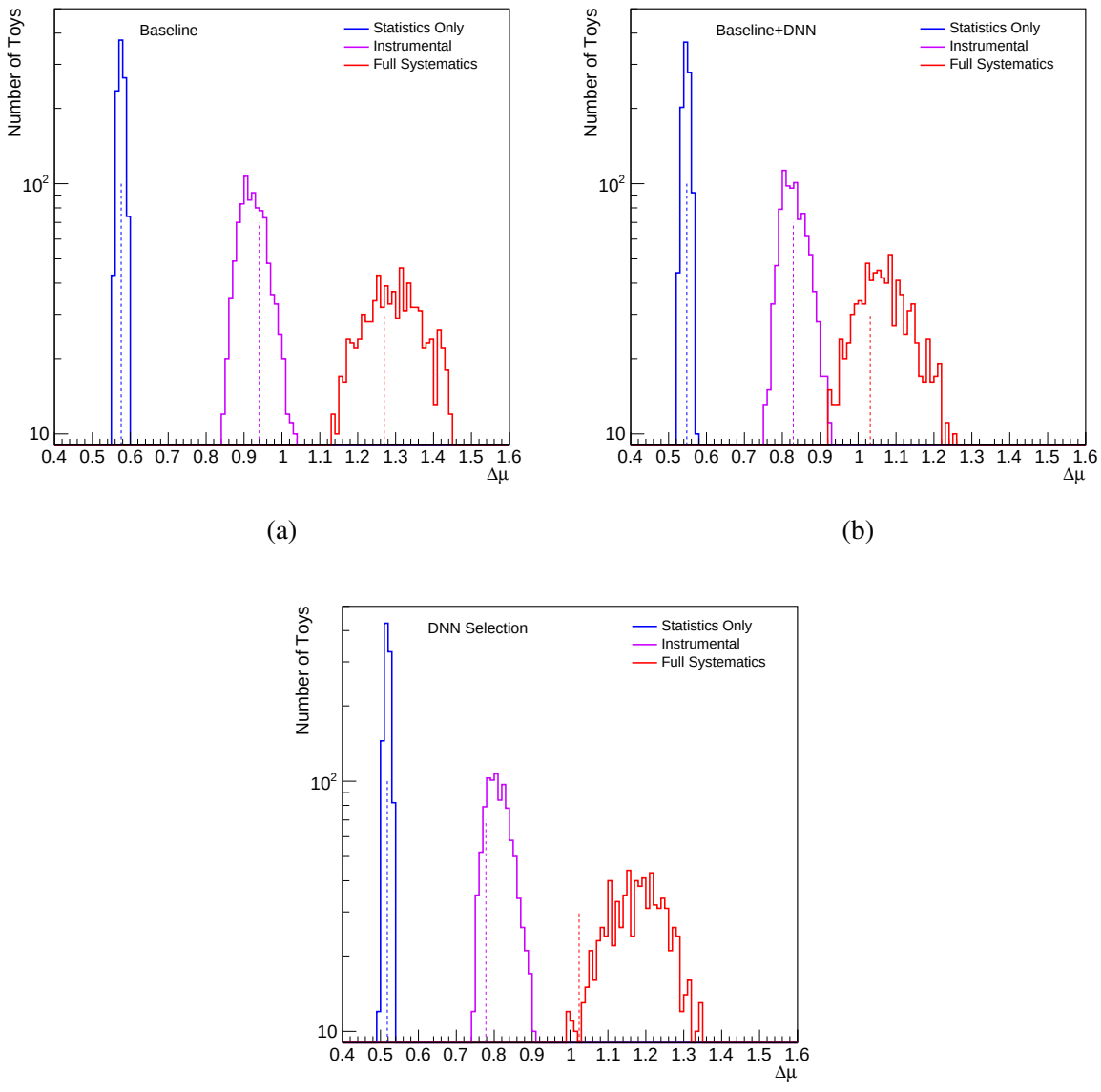


Figure 9.34: The distribution of the error on μ ($\Delta\mu$) in the three systematic models, for (a) the retrained baseline, (b) the baseline+DNN, and (c) the DNN regions. The statistics only fits are shown in blue, the instrumental systematics fits are shown in purple, and the full systematics fits are shown in red. The dashed lines represent the result from the nominal fit without toy weights.

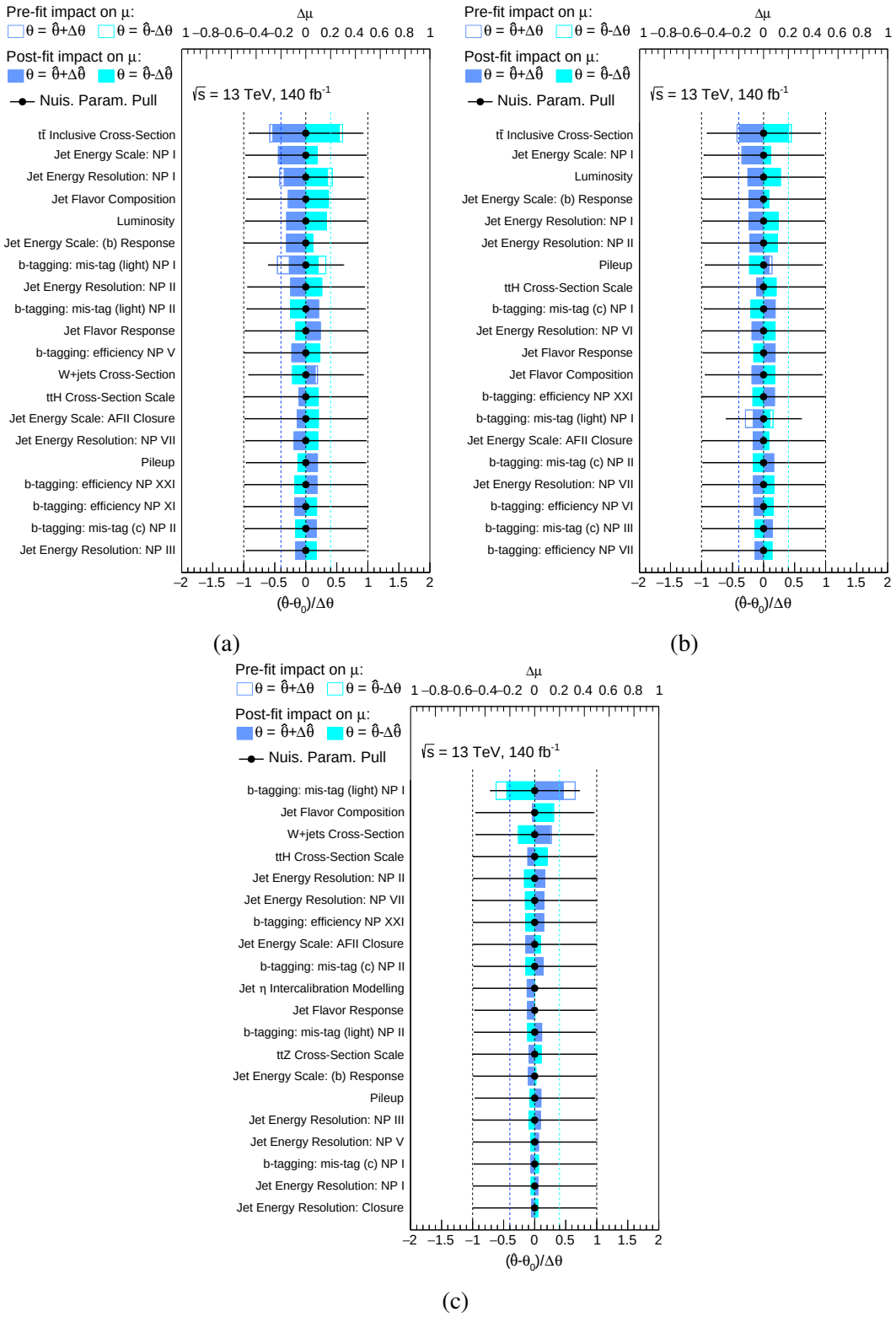


Figure 9.35: The pulls and constraints of the 20 largest systematic uncertainties in the instrumental fits, for (a) the retrained baseline, (b) the baseline+DNN, and (c) the DNN region. The effect of the nuisance parameters on the final result $\Delta\mu$, calculated by comparing the nominal result with the result after fixing the parameter to it's best fit value $\hat{\theta}$, are shown both pre and post fit by the unfilled and filled rectangles respectively. The pulls of the nuisance parameters from the nominal values θ_0 are shown by the deviation of the black points from 0, with 1σ pull marked by dashed lines.

should be approximately equal in the three fit options, and this is confirmed in the fits. It is thus impossible to determine if any improvement from the DNN selection or baseline+DNN BDT is significant. The available MC statistics for the alternative samples used to evaluate the modelling systematics are too small to make any definitive statements on fits including these systematics. Given that these systematics were dominant in the analysis presented in Chapter 8, a new strategy must be sought to overcome this limited. One such strategy is to produce MC samples filtered to include high p_T top quarks, such that the effective statistics in the boosted region is largely enhanced. Should the MC statistics also prove to be problematic in the resolved regions, strategies such as those presented in Section 6.4.1 may also be of use.

9.3.6 DNN Tagging Summary

The statistics only fits demonstrate clearly that there is a nominal improvement by including the DNN into the boosted $t\bar{t}H, H \to b\bar{b}$ analysis, and that both strategies – selecting events using the DNN, or simply using it to create discriminating variables for the classifier – are viable options that improve on the existing baseline. It is important to note that these results alone are not the full story. Combining the boosted region with the resolved regions is imperative to constrain the free floating normalisations of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ backgrounds. Other systematics, such as the light tagging systematics that the DNN selection is more sensitive to, are likely to be constrainable by control regions of the resolved analysis. The DNN based selection also has the advantage of a smaller overlap with the resolved signal regions, meaning it is expected to have a larger relative performance gain in the full combination. However, it is only when there is a full resolved analysis with which to combine that it will become clear which of these regions is the best choice moving forward.

9.4 Outlook

Given that the $t\bar{t}H$ process has now been discovered to $> 5\sigma$ [227], the objectives of the $t\bar{t}H, H \to b\bar{b}$ analysis also evolve. With the full 140 fb^{-1} of data from Run 2, it is possible to begin to measure not only the inclusive $t\bar{t}H$ cross-section, but to investigate differential measurements such as by using the simplified template cross-section framework [234], or the CP structure of the Higgs boson via modifications to the $t\bar{t}H$ vertex [235]. The boosted channel becomes increasingly valuable in these analyses [47], and the improved Higgs reconstruction obtained utilising the DNN described in Section 9.3 is likely to increase the sensitivity of these analyses in the future. On top of this, the jet substructure variables described in Section 9.2 are now available for use in ATLAS analyses. Inclusion of these variables, particularly in the training of the DNN, can only further improve not only the Higgs reconstruction but also the signal and background separation in the classification BDTs. Although the use of track jets, studied in Section 9.1, did not improve the analysis, the twin developments of the jet substructure and the DNN tagger are promising avenues for the boosted channel with 140 fb^{-1}.

Chapter 10

Conclusions

"All you really need to know for the moment is that the universe is a lot more complicated than you might think, even if you start from a position of thinking it's pretty damn complicated in the first place."

– Douglas Adams

This thesis presented two analyses: a measurement of boosted top quark kinematics, and the search for $t\bar{t}H, H \to b\bar{b}$ production with boosted objects, using $\sqrt{s} = 13$ TeV pp collision data recorded with the ATLAS experiment at the LHC.

Using the 3.2 fb^{-1} recorded in 2015, $t\bar{t}$ events were selected in the lepton+jets channel, by requiring a single lepton, a b-tagged small jet, and a top-tagged large jet with $p_T > 300$ GeV. The $p_T^{\text{t,had}}$ and $\left|y^{\text{t,had}}\right|$ spectra of the top-tagged large jet were unfolded to a fiducial phase space, and compared to a number of state-of-the-art MC generators. The rapidity distribution, measured to a precision of approximately 4-10%, was found to be in good agreement with the Standard Model. The p_T distribution, measured with a 5-40% uncertainty that increases with p_T, reconfirms the disagreement seen first in Run 1 at the higher centre-of-mass energy of Run 2.

The presence of $t\bar{t}H$ production, as well as the $H \to b\bar{b}$ decay mode, were established, in part thanks to the analysis of 36 fb^{-1} of data in the $t\bar{t}H, H \to b\bar{b}$ channel. This included, for the first time by ATLAS, a region targeting the high p_T boosted regime. Events were selected in the lepton+jets channel, requiring a single lepton and two reclustered jets of $p_T > 200$ GeV, with subjet b-tagging used to select boosted Higgs boson and top quark candidates. A profile likelihood fit was performed to multivariate distributions, trained to separate signal and background, in each of the signal regions. The sensitivity of the analysis was limited by the modelling of the $t\bar{t}$+≥1b background, which was modelled in MC with a normalisation left free floating in the profile likelihood fit.

New techniques, including the utilisation of deep learning techniques and of novel jet substructure observables, have been developed to further increase sensitivity to these rare and fascinating events in future publications. A multiclass deep neural network is trained on $t\bar{t}H$ MC to tag the origin of high p_T reclustered jets as either a Higgs jet, a top jet, or a QCD jet. This

allows for a new event selection requiring one Higgs jet and one top jet, in which the $S/\sqrt{B}$, Higgs reconstruction efficiency, and overlap with resolved signal regions are all improved relative to the previous analysis. Furthermore, jet substructure variables are defined for reclustered jets for the first time at ATLAS, which show promising performance in discrimination between $t\bar{t}H$ signal and $t\bar{t}$ background.

The LHC is scheduled to turn back on again to collide protons at $\sqrt{s} = 13$ or even 14 TeV in 2021. The new FastTracKer upgrade, currently being installed in ATLAS, is a new hardware upgrade giving earlier access to tracking information in the ATLAS trigger. In this thesis, a new muon trigger chain has been implemented, which incorporates FTK tracks into muon trigger chains. This is now set up and ready to save as much Run 3 data as possible. Strategies have also been developed to generate as much MC as possible in the high luminosity future for interesting signatures in extreme regions of phase space, with example application in the $t\bar{t}H, H \to b\bar{b}$ analysis in which geenrating sufficient MC statistics is a persistent issue.

With data taking planned until at least 2035, and an expected total dataset around twenty times what has thus far been collected, the rich LHC physics program is far from over. Data analysis techniques are developing rapidly, and major discoveries may well be just around the corner.

The Standard Model holds firm – for now.

Bibliography

[1] The ATLAS Collaboration, "Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector", *JHEP* 11:191, 2017, arXiv:1708.00727 [hep-ex]

[2] The ATLAS Collaboration, "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector", *Phys. Rev.* D97(7):072016, 2018, arXiv:1712.08895 [hep-ex]

[3] The ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* 716(1):1 – 29, 2012, ISSN 0370-2693, http://www.sciencedirect.com/science/article/pii/S037026931200857X

[4] The CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* 716(1):30 – 61, 2012, ISSN 0370-2693, http://www.sciencedirect.com/science/article/pii/S0370269312008581

[5] H. YUKAWA, "On the Interaction of Elementary Particles. I", *Nippon Sugaku-Buturigakkwai Kizi Dai 3 Ki* 17:48–57, 1935

[6] D. de Florian et al. (LHC Higgs Cross Section Working Group), "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", 2016, arXiv:1610.07922 [hep-ph]

[7] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$", *Phys.Rev.Lett.* 110:252004, 2013, arXiv:1303.6254 [hep-ph]

[8] D. Hanneke, S. Fogwell, and G. Gabrielse, "New Measurement of the Electron Magnetic Moment and the Fine Structure Constant", *Physical Review Letters* 100(12), March 2008, https://doi.org/10.1103/physrevlett.100.120801

[9] The ATLAS Collaboration, "Standard Model Summary Plots Spring 2019", Technical Report ATL-PHYS-PUB-2019-010, CERN, Geneva, Mar 2019, http://cds.cern.ch/record/2668559

[10] F. Tanedo, "CERN mug summarizes Standard Model, but is off by a factor of 2", June 2011, https://www.quantumdiaries.org/2011/06/26/cern-mug-summarizes-standard-model-but-is-off-by-a-factor-of-2/

[11] Particle Data Group, "Standard Model of Elementary Particles", https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg

[12] E. Noether, "Invariante Variationsprobleme", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918:235–257, 1918, http://eudml.org/doc/59024

[13] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. I", *Phys. Rev. D* 8:3633–3652, Nov 1973, https://link.aps.org/doi/10.1103/PhysRevD.8.3633

[14] D. J. Gross and F. Wilczek, "Asymptotically free gauge theories. II", *Phys. Rev. D* 9:980–993, Feb 1974, https://link.aps.org/doi/10.1103/PhysRevD.9.980

[15] G. M. Prosperi, M. Raciti, and C. Simolo, "On the running coupling constant in QCD", *Prog. Part. Nucl. Phys.* 58:387–438, 2007, arXiv:hep-ph/0607209 [hep-ph]

[16] S. Tomonaga, "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields*", *Progress of Theoretical Physics* 1(2):27–42, 08 1946, ISSN 0033-068X, https://doi.org/10.1143/PTP.1.27

[17] J. Schwinger, "Quantum Electrodynamics. I. A Covariant Formulation", *Phys. Rev.* 74:1439–1461, Nov 1948, https://link.aps.org/doi/10.1103/PhysRev.74.1439

[18] R. P. Feynman, "Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction", *Phys. Rev.* 80:440–457, Nov 1950, https://link.aps.org/doi/10.1103/PhysRev.80.440

[19] T. M. M. Kobayashi, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* 49 (2), 1973

[20] S. W. Herb, et al., "Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions", *Physical Review Letters* 39:252–255, August 1977

[21] M. L. Perl et al., "Evidence for anomalous lepton production in e^+e^- annihilation", *Physical Review Letters* 35:1489–1492, December 1975

[22] The CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys.Rev.Lett.* 74:2626–2631, 1995, arXiv:hep-ex/9503002 [hep-ex]

[23] The D0 Collaboration, "Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.8 TeV", *Phys.Rev.Lett.* 74:2422–2426, 1995, arXiv:hep-ex/9411001 [hep-ex]

[24] The ATLAS Collaboration, "Measurement of the top quark mass in the $t\bar{t} \rightarrow$ lepton+jets channel from $\sqrt{s}$ = 8 TeV ATLAS data and combination with previous results", *Eur. Phys. J.* C79(4):290, 2019, arXiv:1810.01772 [hep-ex]

[25] A. Bettini, *Introduction to Elementary Particle Physics*, 2008, ISBN 9780521880213, http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521880213

[26] S. Moch and P. Uwer, "Theoretical status and prospects for top-quark pair production at hadron colliders", *Phys. Rev.* D78:034003, 2008, arXiv:0804.1476 [hep-ph]

[27] M. Tanabashi et al. (Particle Data Group), "Review of Particle Physics", *Phys. Rev. D* 98:030001, Aug 2018, https://link.aps.org/doi/10.1103/PhysRevD.98.030001

[28] A. Heinson, "Useful Diagrams of Top Signals and Backgrounds", October 2011, https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html

[29] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys.Lett.* 12:132–133, 1964

[30] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* 13:508–509, Oct 1964, http://link.aps.org/doi/10.1103/PhysRevLett.13.508

[31] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* 13:321–323, Aug 1964, http://link.aps.org/doi/10.1103/PhysRevLett.13.321

[32] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* 13:585–587, Nov 1964, http://link.aps.org/doi/10.1103/PhysRevLett.13.585

[33] The ATLAS Collaboration, "Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels with $\sqrt{s}$ = 13 TeV pp collisions using the ATLAS detector", *Phys. Lett.* B784:345–366, 2018, arXiv:1806.00242 [hep-ex]

[34] J. Ellis, "Higgs Physics", "Proceedings, 2013 European School of High-Energy Physics (ESHEP 2013): Paradfurdo, Hungary, June 5-18, 2013", 117–168, 2015, arXiv:1312.5672 [hep-ph]

[35] The ATLAS and CMS Collaborations, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}$ = 7 and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.* 114:191803, 2015, arXiv:1503.07589 [hep-ex]

[36] J. R. Andersen et al. (LHC Higgs Cross Section Working Group), "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", 2013, arXiv:1307.1347 [hep-ph]

[37] A. Randle-Conde, "What next for the Higgs?", June 2012, https://www.quantumdiaries.org/2012/06/09/what-next-for-the-higgs/

[38] The ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV", *JHEP* 08:045, 2016, arXiv:1606.02266 [hep-ex]

[39] The CMS Collaboration, "Measurement of the top quark Yukawa coupling from $\mathrm{t\bar{t}}$ kinematic distributions in the lepton+jets final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV", 2019, arXiv:1907.01590 [hep-ex]

[40] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, "Associated production of Higgs and single top at hadron colliders", *Phys. Rev.* D64:094023, 2001, arXiv:hep-ph/0106293 [hep-ph]

[41] The ATLAS Collaboration, "Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector", *JHEP* 05:160, 2016, arXiv:1604.03812 [hep-ex]

[42] The ATLAS Collaboration, "Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector", *Phys. Lett.* B749:519–541, 2015, arXiv:1506.05988 [hep-ex]

[43] The ATLAS Collaboration, "Search for $H \rightarrow \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector", *Phys.Lett.* B740:222–242, 2015, arXiv:1409.3122 [hep-ex]

[44] The CMS Collaboration, "Search for a standard model Higgs boson produced in association with a top-quark pair and decaying to bottom quarks using a matrix element method", *CMS-HIG-14-010, CERN-PH-EP-2015-016* 2015, arXiv:1502.02485 [hep-ex]

[45] The CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* 09:087, 2014, [Erratum: JHEP10,106(2014)], arXiv:1408.1682 [hep-ex]

[46] S. Dittmaier et al. (LHC Higgs Cross Section Working Group), "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", 2011, arXiv:1101.0593 [hep-ph]

[47] N. Moretti, P. Petrov, S. Pozzorini, and M. Spannowsky, "Measuring the signal strength in $t\bar{t}H$ with $H \to b\bar{b}$", *Phys. Rev.* D93(1):014019, 2016, arXiv:1510.08468 [hep-ph]

[48] G. F. Giudice, "Naturalness after LHC8", *PoS* EPS-HEP2013:163, 2013, arXiv:1307.7879 [hep-ph]

[49] S. P. Martin, "A Supersymmetry primer", 1–98, 1997, [Adv. Ser. Direct. High Energy Phys.18,1(1998)], arXiv:hep-ph/9709356 [hep-ph]

[50] J.-F. Arguin, "Experimental interplay between the top quark and Supersymmetry at the LHC and Tevatron", *Journal of Physics* 452 012005, 2013

[51] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria, "Handbook of vectorlike quarks: Mixing and single production", *Phys. Rev.* D88(9):094010, 2013, arXiv:1306.0572 [hep-ph]

[52] Q.-H. Cao, B. Yan, J.-H. Yu, and C. Zhang, "A General Analysis of Wtb anomalous Couplings", *Chin. Phys.* C41(6):063101, 2017, arXiv:1504.03785 [hep-ph]

[53] J. A. Aguilar-Saavedra, "A minimal set of top anomalous couplings", *Nucl. Phys.* B812:181–204, 2009, arXiv:0811.3842 [hep-ph]

[54] C. Degrande, J.-M. Gerard, C. Grojean, F. Maltoni, and G. Servant, "Non-resonant New Physics in Top Pair Production at Hadron Colliders", *JHEP* 03:125, 2011, arXiv:1010.6304 [hep-ph]

[55] N. Greiner, S. Willenbrock, and C. Zhang, "Effective Field Theory for Nonstandard Top Quark Couplings", *Phys. Lett.* B704:218–222, 2011, arXiv:1104.3122 [hep-ph]

[56] A. Nelson, "Naturally weak CP violation", *Physics Letters B* 136(5-6):387–391, March 1984, https://doi.org/10.1016/0370-2693(84)92025-2

[57] C. Degrande, J. M. Gerard, C. Grojean, F. Maltoni, and G. Servant, "Probing Top-Higgs Non-Standard Interactions at the LHC", *JHEP* 07:036, 2012, [Erratum: JHEP03,032(2013)], arXiv:1205.1065 [hep-ph]

[58] F. Maltoni, E. Vryonidou, and C. Zhang, "Higgs production in association with a top-antitop pair in the Standard Model Effective Field Theory at NLO in QCD", *JHEP* 10:123, 2016, arXiv:1607.05330 [hep-ph]

[59] D. Buttazzo et al., "Investigating the near-criticality of the Higgs boson", *JHEP* 12:089, 2013, arXiv:1307.3536 [hep-ph]

[60] O. Behnke, K. Kroeninger, T. Schoerner-Sadenius, and G. Schott, *Data Analysis in High Energy Physics*, Wiley, 2013

[61] A. Armbruster, K. Kroeninger, B. Malaescu, and F. Spano, "Practical considerations for unfolding", Technical Report ATL-COM-PHYS-2014-277, CERN, Geneva, Apr 2014, https://cds.cern.ch/record/1694351

[62] G. Choudalakis, "Fully Bayesian Unfolding", 2012, arXiv:1201.4612 [physics.data-an]

[63] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding", *Nucl. Instrum. Meth.* A372:469–481, 1996, arXiv:hep-ph/9509307 [hep-ph]

[64] G. D'Agostini, "A Multidimensional Unfolding Method Based On Bayes' Theorem", *Nucl. Instr. and Meth. in Phys. Res.* A362:487–498, 1995

[65] T. Adye, "Unfolding algorithms and tests using RooUnfold", "Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011, CERN-2011-006, pp 313-318", 313–318, 2011, arXiv:1105.1160 [physics.data-an]

[66] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf* C0303241:MOLT007, 2003, [,186(2003)], arXiv:physics/0306116 [physics]

[67] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* C71:1554, 2011, [Erratum: Eur. Phys. J.C73,2501(2013)], arXiv:1007.1727 [physics.data-an]

[68] O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius (Editors), *Data Analysis in High Energy Physics*, Wiley-VCH Verlag GmbH & Co. KGaA, June 2013, https://doi.org/10.1002/9783527653416

[69] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", *Journal of Computer and System Sciences* 55(1):119–139, August 1997, https://doi.org/10.1006/jcss.1997.1504

[70] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature Cell Biology* 521(7553):436–444, 5 2015, ISSN 1465-7392

[71] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, "Weakly Supervised Classification in High Energy Physics", *JHEP* 05:145, 2017, arXiv:1702.00414 [hep-ph]

[72] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, "JUNIPR: a Framework for Unsupervised Machine Learning in Particle Physics", *Eur. Phys. J.* C79(2):102, 2019, arXiv:1804.09720 [hep-ph]

[73] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees. Wadsworth Int", *Group* 37(15):237–251, 1984

[74] A. Hoecker et al., "TMVA: Toolkit for Multivariate Data Analysis", *PoS* ACAT:040, 2007, arXiv:physics/0703039

[75] C. Gini, "Variabilità e mutabilità (Variability and Mutability)", *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955) ed. Bologna* 1912

[76] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A Tutorial on the Cross-Entropy Method", *Annals of Operations Research* 134(1):19–67, Feb 2005, ISSN 1572-9338, https://doi.org/10.1007/s10479-005-5724-z

[77] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine", *Annals of Statistics* 29:1189–1232, 2000

[78] C. Asawa, "Stanford University CS231N Lecture Materials", 2019, http://cs231n.github.io/neural-networks-1/

[79] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back Propagating Errors", *Nature* 323:533–536, 10 1986

[80] S. Miller, "How to Build a Neural Network", 2015, https://stevenmiller888.github.io/mind-how-to-build-a-neural-network/

[81] D. Nam, H. Kim, J. Cho, and R. Jayakrishnan, "A Model Based on Deep Learning for Predicting Travel Mode Choice", 01 2017, https://www.researchgate.net/publication/317913178_A_Model_Based_on_Deep_Learning_for_Predicting_Travel_Mode_Choice

[82] E. L. Lehmann and G. Casella, *Theory of point estimation*, Springer Texts in Statistics, Springer-Verlag, New York, second edition, 1998, ISBN 0-387-98502-6

[83] Y. LeCun et al., "Backpropagation Applied to Handwritten Zip Code Recognition", *Neural Comput.* 1(4):541–551, December 1989, ISSN 0899-7667, http://dx.doi.org/10.1162/neco.1989.1.4.541

[84] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation* 9(8):1735–1780, nov 1997, https://doi.org/10.1162/neco.1997.9.8.1735

[85] I. Goodfellow et al., "Generative Adversarial Nets", Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Editors), "Advances in Neural Information Processing Systems 27", 2672–2680, Curran Associates, Inc., 2014, http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[86] L. Evans and P. Bryant, "LHC Machine", *JINST* 3:S08001, 2008

[87] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* 3(08):S08003–S08003, aug 2008, https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003

[88] The CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* 3:S08004, 2008

[89] The LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* 3:S08005, 2008

[90] The ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* 3:S08002, 2008

[91] S. Myers and E. Picasso, "The design, construction and commissioning of the CERN large Electron–Positron collider", *Contemporary Physics* 31(6):387–403, 1990, https://doi.org/10.1080/00107519008213789, https://doi.org/10.1080/00107519008213789

[92] G. Arnison et al. (UA1), "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV", *Physics Letters B* 122(1):103 – 116, 1983, ISSN 0370-2693, http://www.sciencedirect.com/science/article/pii/0370269383911772

[93] M. Banner et al., "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider", *Physics Letters B* 122(5):476 – 485, 1983, ISSN 0370-2693, http://www.sciencedirect.com/science/article/pii/0370269383916052

[94] G. Arnison et al. (UA1), "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider", *Phys. Lett.* B126:398–410, 1983, [,7.55(1983)]

[95] P. Bagnaia et al. (UA2), "Evidence for Z0 —> e+ e- at the CERN anti-p p Collider", *Phys. Lett.* B129:130–140, 1983, [,7.69(1983)]

[96] Science and Technologies Funding Council, "CERN Accelerator Complex", 2016, https://stfc.ukri.org/research/particle-physics-and-particle-astrophysics/large-hadron-collider/cern-accelerator-complex/

[97] P. Grafström and W. Kozanecki, "Luminosity determination at proton colliders", *Progress in Particle and Nuclear Physics* 81:97 – 148, 2015, ISSN 0146-6410, http://www.sciencedirect.com/science/article/pii/S0146641014000878

[98] J. Wenninger, "Operation and Configuration of the LHC in Run 2", Mar 2019, https://cds.cern.ch/record/2668326

[99] The ATLAS Collaboration, "Luminosity Summary Plots", 2019, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2

[100] A. G. et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, CERN, Geneva, 2017, https://cds.cern.ch/record/2284929

[101] J. Pequenao and P. Schaffner, "How ATLAS detects particles: diagram of particle paths in the detector", Jan 2013, https://cds.cern.ch/record/1505342

[102] *ATLAS inner detector: Technical Design Report, 1*, Technical Design Report ATLAS, CERN, Geneva, 1997, https://cds.cern.ch/record/331063

[103] J. Pequenao, "Computer generated image of the ATLAS inner detector", Mar 2008, https://cds.cern.ch/record/1095926

[104] M. S. Alam et al. (ATLAS), "ATLAS pixel detector: Technical design report", 1998

[105] S. Hawking, *A Brief History of Time*, Bantam Books, 1998, ISBN 9780553109535, https://books.google.com/books?id=E7mEnx3zE4AC

[106] M. Capeans et al. (ATLAS Collaboration), "ATLAS Insertable B-Layer Technical Design Report", Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010, https://cds.cern.ch/record/1291633

[107] Y. Unno, "ATLAS silicon microstrip detector system (SCT)", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 511(1):58 – 63, 2003, ISSN 0168-9002, proceedings of the 11th International Workshop on Vertex Detectors, http://www.sciencedirect.com/science/article/pii/S0168900203017510

[108] T. A. T. collaboration et al., "The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance", *Journal of Instrumentation* 3(02):P02013–P02013, feb 2008, https://doi.org/10.1088%2F1748-0221%2F3%2F02%2Fp02013

[109] A. Airapetian et al. (ATLAS), "ATLAS calorimeter performance Technical Design Report", 1996

[110] J. Pequenao, "Computer Generated image of the ATLAS calorimeter", Mar 2008, https://cds.cern.ch/record/1095927

[111] *ATLAS liquid-argon calorimeter: Technical Design Report*, Technical Design Report ATLAS, CERN, Geneva, 1996, https://cds.cern.ch/record/331061

[112] *ATLAS tile calorimeter: Technical Design Report*, Technical Design Report ATLAS, CERN, Geneva, 1996, https://cds.cern.ch/record/331062

[113] *ATLAS muon spectrometer: Technical Design Report*, Technical Design Report ATLAS, CERN, Geneva, 1997, https://cds.cern.ch/record/331068

[114] J. Pequenao, "Computer generated image of the ATLAS Muons subsystem", Mar 2008, https://cds.cern.ch/record/1095929

[115] The ATLAS Collaboration, "Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s}$ = 8 TeV", *Eur. Phys. J.* C75:120, 2015, arXiv:1408.3179 [hep-ex]

[116] M. Livan (ATLAS Muon Group), "Monitored drift tubes in ATLAS", Technical Report ATL-M-PN-129, CERN, Geneva, Sep 1996, https://cds.cern.ch/record/319197

[117] T. Argyropoulos et al., "Cathode Strip Chambers in ATLAS: Installation, Commissioning and in Situ Performance", *Nuclear Science, IEEE Transactions on* 56:1568 – 1574, 07 2009

[118] G. C. and, "The Resistive Plate Chambers of the ATLAS experiment: performance studies", *Journal of Physics: Conference Series* 280:012001, feb 2011, https://doi.org/10.1088%2F1742-6596%2F280%2F1%2F012001

[119] "ATLAS magnet system: Technical design report", 1997

[120] A. Borga et al., "The C-RORC PCIe card and its application in the ALICE and ATLAS experiments", *Journal of Instrumentation* 10:C02022–C02022, 02 2015

[121] "ATLAS first level trigger: Technical design report", 1998

[122] P. Jenni, M. Nessi, M. Nordberg, and K. Smith (ATLAS Collaboration), *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*, Technical Design Report ATLAS, CERN, Geneva, 2003, https://cds.cern.ch/record/616089

[123] M. Shochet et al., "Fast TracKer (FTK) Technical Design Report", Technical Report CERN-LHCC-2013-007. ATLAS-TDR-021, CERN, Geneva, Jun 2013, aTLAS Fast Tracker Technical Design Report, http://cds.cern.ch/record/1552953

[124] The ATLAS Collaboration, "Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data", *Eur. Phys. J. C* (arXiv:1404.2240. CERN-PH-EP-2014-040):74. 38 p, Apr 2014, http://cds.cern.ch/record/1694142

[125] W. Lampl et al., "Calorimeter Clustering Algorithms: Description and Performance", Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008, http://cds.cern.ch/record/1099735

[126] B Aubert et al., "Performance of the ATLAS electromagnetic calorimeter end-cap module 0", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 500(1):178 – 201, 2003, ISSN 0168-9002, nIMA Vol 500, http://www.sciencedirect.com/science/article/pii/S0168900203003449

[127] The ATLAS Collaboration, "Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC", *Eur. Phys. J.* C73(3):2305, 2013, arXiv:1203.1302 [hep-ex]

[128] A. Rosenfeld and J. L. Pfaltz, "Sequential Operations in Digital Picture Processing", *J. ACM* 13(4):471–494, October 1966, ISSN 0004-5411, http://doi.acm.org/10.1145/321356.321357

[129] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME–Journal of Basic Engineering* 82(Series D):35–45, 1960

[130] The ATLAS Collaboration, "A neural network clustering algorithm for the ATLAS silicon pixel detector", *JINST* 9:P09009, 2014, arXiv:1406.7690 [hep-ex]

[131] The ATLAS Collaboration, "Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2", *Eur. Phys. J.* C77(10):673, 2017, arXiv:1704.07983 [hep-ex]

[132] The ATLAS Collaboration, "Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s}$ = 13 TeV", Technical Report ATL-PHYS-PUB-2015-051, CERN, Geneva, Dec 2015, https://cds.cern.ch/record/2110140

[133] The ATLAS Collaboration, "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC", *Eur. Phys. J.* C77(5):332, 2017, arXiv:1611.10235 [physics.ins-det]

[134] The ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV", *Eur. Phys. J.* C76(5):292, 2016, arXiv:1603.05598 [hep-ex]

[135] J. Illingworth and J. Kittler, ""A survey of the Hough transform"", *Computer Vision, Graphics, and Image Processing* 44(1):87–116, October 1988, https://doi.org/10.1016/s0734-189x(88)80033-1

[136] The ATLAS Collaboration, "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV", *Submitted to: Eur. Phys. J.* 2019, arXiv:1902.04655 [physics.ins-det]

[137] The ATLAS Collaboration, "Electron and photon energy calibration with the ATLAS detector using 2015-2016 LHC proton-proton collision data", *Submitted to: JINST* 2018, arXiv:1812.03848 [hep-ex]

[138] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Physics Letters B* 659(1-2):119–126, jan 2008, https://doi.org/10.1016/j.physletb.2007.09.077

[139] J. E. Huth et al., "Toward a standardization of jet definitions", "1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990", 0134–136, 1990, http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-249

[140] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions", *Nucl. Phys.* B406:187–224, 1993

[141] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP* 08:001, 1997, arXiv:hep-ph/9707323 [hep-ph]

[142] M. Cacciari, G. Salam, and G. Soyez, "The anti-k_T jet clustering algorithm", *JHEP 0804:063,2008*

[143] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J.* C72:1896, 2012, arXiv:1111.6097 [hep-ph]

[144] The ATLAS Collaboration, "Performance of pile-up mitigation techniques for jets in *pp* collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector", *Eur. Phys. J.* C76(11):581, 2016, arXiv:1510.03823 [hep-ex]

[145] The ATLAS Collaboration, "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector", *Phys. Rev.* D96(7):072002, 2017, arXiv:1703.09665 [hep-ex]

[146] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* 04:005, 2008, arXiv:0802.1188 [hep-ph]

[147] The ATLAS Collaboration, "Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS", Technical Report ATLAS-CONF-2015-037, CERN, Geneva, Aug 2015, https://cds.cern.ch/record/2044941

[148] The ATLAS Collaboration, "Expected performance of the ATLAS *b*-tagging algorithms in Run-2", Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015, http://cds.cern.ch/record/2037697

[149] The ATLAS Collaboration (ATLAS Collaboration), "Optimisation of the ATLAS *b*-tagging performance for the 2016 LHC Run", Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun 2016, https://cds.cern.ch/record/2160731

[150] The ATLAS Collaboration, "Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s}$ = 13 TeV", *JHEP* 08:089, 2018, arXiv:1805.01845 [hep-ex]

[151] N. Bartosik, "B-tagging Diagram", May 2016, https://en.wikipedia.org/wiki/File:B-tagging_diagram.png

[152] The ATLAS Collaboration, "Measurement of *b*-tagging Efficiency of *c*-jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector", Technical Report ATLAS-CONF-2018-001, CERN, Geneva, Mar 2018, https://cds.cern.ch/record/2306649

[153] The ATLAS Collaboration, "Calibration of the *b*-tagging efficiency on charm jets using a sample of W+c events with $\sqrt{s}$ = 13 TeV ATLAS data", Technical Report ATLAS-CONF-2018-055, CERN, Geneva, Dec 2018, https://cds.cern.ch/record/2652195

[154] The ATLAS Collaboration, "Calibration of light-flavour jet *b*-tagging rates on ATLAS proton-proton collision data at $\sqrt{s}$ = 13 TeV", Technical Report ATLAS-CONF-2018-006, CERN, Geneva, Apr 2018, https://cds.cern.ch/record/2314418

[155] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming", *JHEP* 02:084, 2010, arXiv:0912.1342 [hep-ph]

[156] The ATLAS Collaboration, "Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector", *Eur. Phys. J.* C75:17, 2015, arXiv:1406.0076 [hep-ex]

[157] The ATLAS Collaboration, "Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector", *JHEP* 09:076, 2013, arXiv:1306.4945 [hep-ex]

[158] Torbjörn Sjöstrand et al., "An Introduction to Pythia 8.2", (CERN-PH-TH-2014-190), 2014

[159] S. Marzani, G. Soyez, and M. Spannowsky, "Looking inside jets: an introduction to jet substructure and boosted-object phenomenology", 2019, [Lect. Notes Phys.958,pp.(2019)], arXiv:1901.10342 [hep-ph]

[160] S. Catani, Y. Dokshitzer, M. Seymour, and B. Webber, "Longitudinally-invariant k⊥-clustering algorithms for hadron-hadron collisions", *Nuclear Physics B* 406(1-2):187–224, September 1993, https://doi.org/10.1016/0550-3213(93)90166-m

[161] J. Thaler and K. Van Tilburg, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", *JHEP* 02:093, 2012, arXiv:1108.2701 [hep-ph]

[162] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure", *JHEP* 06:108, 2013, arXiv:1305.0007 [hep-ph]

[163] The ATLAS Collaboration, "Boosted hadronic top identification at ATLAS for early 13 TeV data", (ATL-PHYS-PUB-2015-053), Dec 2015, https://cds.cern.ch/record/2116351

[164] T. Plehn and M. Spannowsky, "Top Tagging", *J. Phys.* G39:083001, 2012, arXiv:1112.4441 [hep-ph]

[165] The ATLAS Collaboration, "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV", *Eur. Phys. J.* C78(11):903, 2018, arXiv:1802.08168 [hep-ex]

[166] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and Collider Physics", *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 8:1–435, 1996

[167] F. Aaron et al. (H1, ZEUS), "Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA", *JHEP* 1001:109, 2010, arXiv:0911.0884 [hep-ex]

[168] Z. Collaboration", *The ZEUS Detector Technical Proposal*, DESY-HERA-ZEUS-1, 1986, https://cds.cern.ch/record/1478623

[169] H1 Collaboration, *Technical proposal for the H1 detector*, DESY, Hamburg, 1986, https://cds.cern.ch/record/108544

[170] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera, and J. Rojo, "Pinning down the large-x gluon with NNLO top-quark pair differential distributions", *JHEP* 04:044, 2017, arXiv:1611.08609 [hep-ph]

[171] G. R. Stefano Frixione, Paolo Nason, "The POWHEG-hvq manual version 1.0", (Bicocca-FT-07-10, GEF-TH-16/2007), July 2007

[172] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* 1407:079, 2014, arXiv:1405.0301 [hep-ph]

[173] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* 02:007, 2009, arXiv:0811.4622 [hep-ph]

[174] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys.* B126:298–318, 1977

[175] S. M. T. Sjöstrand and P. Z. Skands, "Pythia6.4 Manual", (0605 JHEP 026 [hep-ph/0603175]), 2006

[176] L. Lönnblad, "Fooling Around with the Sudakov Veto Algorithm", *Eur. Phys. J.* C73(3):2350, 2013, arXiv:1211.7204 [hep-ph]

[177] A. Buckley et al., "Rivet user manual", *Comput. Phys. Commun.* 184:2803–2819, 2013, arXiv:1003.0694 [hep-ph]

[178] The ATLAS Collaboration, "Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector", *Eur. Phys. J.* C76(1):11, 2016, arXiv:1508.06868 [hep-ex]

[179] G. I. B. Andersson, G. Gustafson and T. Sjöstrand, "Parton fragmentation and string dynamics", *Phys. Rept.* 97:31–145, 1983

[180] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J.* C58:639–707, 2008, arXiv:0803.0883 [hep-ph]

[181] The ATLAS Collaboration, "Measurement of charged-particle distributions sensitive to the underlying event in $\sqrt{s}$ = 13 TeV proton-proton collisions with the ATLAS detector at the LHC", *JHEP* 03:157, 2017, arXiv:1701.05390 [hep-ex]

[182] S. Agostinelli et al., "Geant4 — A Simulation Toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506(3):250 – 303, 2003, ISSN 0168-9002, http://www.sciencedirect.com/science/article/pii/S0168900203013688

[183] W. Lukas, "Fast Simulation for ATLAS: Atlfast-II and ISF", 2012, http://cdsweb.cern.ch/record/1458503/files/ATL-SOFT-PROC-2012-065.pdf

[184] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* 06:043, 2010, arXiv:1002.2581 [hep-ph]

[185] H.-L. Lai et al., "New parton distributions for collider physics", *Phys.Rev.* D82:074024, 2010, arXiv:1007.2241 [hep-ph]

[186] D. Stump et al., "Inclusive jet production, parton distributions, and the search for new physics", *JHEP* 10:046, 2003, arXiv:hep-ph/0303013 [hep-ph]

[187] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys.Rev.* D82:074018, 2010, arXiv:1005.3457 [hep-ph]

[188] The ATLAS Collaboration, "Comparison of Monte Carlo generator predictions from Powheg and Sherpa to ATLAS measurements of top pair production at 7 TeV", Technical Report ATL-PHYS-PUB-2015-011, CERN, Geneva, May 2015, https://cds.cern.ch/record/2020602

[189] The ATLAS Collaboration, "ATLAS Run 1 Pythia8 tunes", Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014, https://cds.cern.ch/record/1966419

[190] R. D. Ball et al. (The NNPDF Collaboration), "Parton distributions for the LHC Run II", *EDINBURGH-2014-15, IFUM-1034-FT, CERN-PH-TH-2013-253, OUTP-14-11P, CAVENDISH-HEP-14-11* 2014, arXiv:1410.8849 [hep-ph]

[191] J. Bellm et al., "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J. C* 76(4):196, 2016, arXiv:1512.01178 [hep-ph]

[192] F. Cascioli, P. Maierhofer, and S. Pozzorini, "Scattering Amplitudes with Open Loops", *Phys. Rev. Lett.* 108:111601, 2012, arXiv:1111.5206 [hep-ph]

[193] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* 185:2930, 2014, arXiv:1112.5675 [hep-ph]

[194] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* 03:015, 2013, arXiv:1212.3460 [hep-ph]

[195] S. Alekhin, J. Blumlein, S. Klein, and S. Moch, "The 3, 4, and 5-flavor NNLO Parton from Deep-Inelastic-Scattering Data and at Hadron Colliders", *Phys. Rev.* D81:014032, 2010, arXiv:0908.2766 [hep-ph]

[196] H.-L. Lai et al., "New parton distributions for collider physics", *Phys. Rev.* D82:074024, 2010, arXiv:1007.2241 [hep-ph]

[197] The ATLAS Collaboration, "Studies of tt+cc production with MadGraph5_aMC@NLO and Herwig++ for the ATLAS experiment", Technical Report ATL-PHYS-PUB-2016-011, CERN, Geneva, May 2016, https://cds.cern.ch/record/2153876

[198] R. Raitio and W. W. Wada, "Higgs-boson production at large transverse momentum in quantum chromodynamics", *Phys. Rev. D* 19:941–944, Feb 1979, https://link.aps.org/doi/10.1103/PhysRevD.19.941

[199] W. Beenakker et al., "NLO QCD corrections to t anti-t H production in hadron collisions", *Nucl. Phys.* B653:151–203, 2003, arXiv:hep-ph/0211352 [hep-ph]

[200] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackeroth, "Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections", *Phys. Rev.* D68:034022, 2003, arXiv:hep-ph/0305087 [hep-ph]

[201] Y. Zhang, W.-G. Ma, R.-Y. Zhang, C. Chen, and L. Guo, "QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider", *Phys. Lett.* B738:1–5, 2014, arXiv:1407.1110 [hep-ph]

[202] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, "Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons", *JHEP* 06:184, 2015, arXiv:1504.03446 [hep-ph]

[203] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension", *Comput. Phys. Commun.* 108:56–74, 1998, arXiv:hep-ph/9704448 [hep-ph]

[204] N. Kidonakis, "NNLL resummation for s-channel single top quark production", *Phys. Rev.* D81:054028, 2010, arXiv:1001.5034 [hep-ph]

[205] N. Kidonakis, "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production", *Phys. Rev.* D83:091503, 2011, arXiv:1103.2792 [hep-ph]

[206] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-", *Phys. Rev.* D82:054018, 2010, arXiv:1005.4451 [hep-ph]

[207] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, "Single-top hadroproduction in association with a W boson", *JHEP* 07:029, 2008, arXiv:0805.3067 [hep-ph]

[208] T. Gleisberg and S. Höche, "Comix, a new matrix element generator", *Journal of High Energy Physics* 2008(12):039–039, dec 2008, https://doi.org/10.1088%2F1126-6708%2F2008%2F12%2F039

[209] The ATLAS Collaboration, "Measurement of W and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s}$ = 13 TeV in the ATLAS Detector", Technical Report ATLAS-CONF-2015-039, CERN, Geneva, Aug 2015, https://cds.cern.ch/record/2045487

[210] The ATLAS Collaboration, "Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector", *Eur. Phys. J.* C76(10):538, 2016, arXiv:1511.04716 [hep-ex]

[211] The ATLAS Collaboration, "Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s}$ = 8 TeV proton-proton collisions using the ATLAS detector", *Phys. Rev.* D93(3):032009, 2016, arXiv:1510.03818 [hep-ex]

[212] The CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s}$ = 8 TeV", *Eur. Phys. J.* C75(11):542, 2015, arXiv:1505.04480 [hep-ex]

[213] The CMS Collaboration, "Measurement of the integrated and differential t-tbar production cross sections for high-pt top quarks in pp collisions at sqrt(s) = 8 TeV", *Phys. Rev.* D94(7):072002, 2016, arXiv:1605.00116 [hep-ex]

[214] A. D. Martin et al., "Parton distributions for the LHC.", *Eur. Phys. J.* C63:189–285, 2009

[215] The ATLAS Collaboration, "A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector", *JHEP* 08:148, 2015, arXiv:1505.07018 [hep-ex]

[216] The ATLAS Collaboration, "Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector", (ATLAS-CONF-2014-058), Oct 2014, http://cds.cern.ch/record/1951336

[217] M. Whalley, D. Bourilkov, and R. Group, "The Les Houches accord PDFs (LHAPDF) and LHAGLUE", 2005, arXiv:hep-ph/0508110 [hep-ph]

[218] E. Maguire, L. Heinrich, and G. Watt, "HEPData: a repository for high energy physics data", *J. Phys. Conf. Ser.* 898(10):102006, 2017, arXiv:1704.05473 [hep-ex]

[219] The ATLAS Collaboration, "In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector", *Eur. Phys. J.* C79(2):135, 2019, arXiv:1807.09477 [hep-ex]

[220] The ATLAS Collaboration, "Improvements in $t\bar{t}$ modelling using NLO+PS Monte Carlo generators for Run2", Technical Report ATL-PHYS-PUB-2018-009, CERN, Geneva, Jul 2018, http://cds.cern.ch/record/2630327

[221] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$", *Phys. Rev. Lett.* 110:252004, Jun 2013, http://link.aps.org/doi/10.1103/PhysRevLett.110.252004

[222] P.-F. Duan, Y. Zhang, Y. Wang, M. Song, and G. Li, "Electroweak corrections to top quark pair production in association with a hard photon at hadron colliders", *Phys. Lett.* B766:102–106, 2017, arXiv:1612.00248 [hep-ph]

[223] L. Moore et al., "TopFitter: A global fit of the top effective theory to data", *PoS* TOP2015:065, 2016

[224] S. Brown et al., "TopFitter: Fitting top-quark Wilson Coefficients to Run II data", "39th International Conference on High Energy Physics (ICHEP 2018) Seoul, Korea, July 4-11, 2018", 2019, arXiv:1901.03164 [hep-ph]

[225] N. P. Hartland et al., "A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector", *JHEP* 04:100, 2019, arXiv:1901.05965 [hep-ph]

[226] The ATLAS Collaboration, "Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector", *Phys. Rev.* D97(7):072003, 2018, arXiv:1712.08891 [hep-ex]

[227] The ATLAS Collaboration, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector", *Phys. Lett.* B784:173–191, 2018, arXiv:1806.00425 [hep-ex]

[228] The ATLAS Collaboration, "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector", *Phys. Lett.* B786:59–86, 2018, arXiv:1808.08238 [hep-ex]

[229] L. Asquith et al., "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in boosted topologies at $\sqrt{s}$=13TeV with the ATLAS detector.", Technical Report ATL-COM-PHYS-2017-396, CERN, Geneva, Apr 2017, https://cds.cern.ch/record/2260232

[230] H. W. Kuhn, "The Hungarian method for the assignment problem", *Naval Research Logistics Quarterly* 2(1-2):83–97, March 1955, https://doi.org/10.1002/nav.3800020109

[231] F. Chollet et al., "Keras", https://keras.io, 2015

[232] Martín Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", 2015, software available from tensorflow.org, http://tensorflow.org/

[233] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *CoRR* 2016, arXiv:1603.02754

[234] M. Delmastro et al., "Simplified Template Cross Sections – Stage 1.1", Technical Report arXiv:1906.02754, CERN, Geneva, Apr 2019, 14 pages, 3 figures, https://cds.cern.ch/record/2669925

[235] D. Azevedo, A. Onofre, F. Filthaut, and R. Gonçalo, "CP tests of Higgs couplings in $t\bar{t}h$ semileptonic events at the LHC", *Phys. Rev.* D98(3):033004, 2018, arXiv:1711.05292 [hep-ph]